Catalogue no. 75F0002MIE — No. 008
ISSN: 1707-2840
ISBN: 0-662-37553-X

Research Paper

Income research paper series

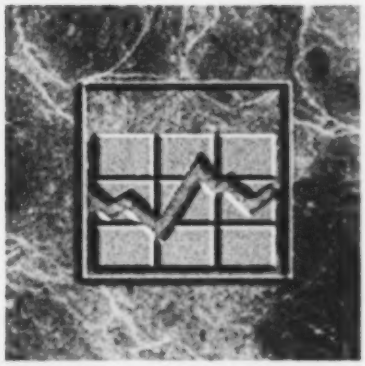
Combined-panel longitudinal weighting Survey of Labour and Income Dynamics

1996-2002

by Jean-François Naud

Income Statistics Division
Jean Talon Building, Ottawa, K1A 0T6

Telephone: 613 951-7355



This paper represents the views of the author and does not necessarily reflect the opinions of Statistics Canada.

Statistics Canada Statistique Canada

Canada

106-04266/

How to obtain more information

Specific inquiries about this product and related statistics or services should be directed to Client Services, Income Statistics Division, Statistics Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355; income@statcan.ca).

For information on the wide range of data available from Statistics Canada, you can contact us by calling one of our toll-free numbers. You can also contact us by e-mail or by visiting our Web site.

National inquiries line	**1 800 263-1136**
National telecommunications device for the hearing impaired	**1 800 363-7629**
Depository Services Program inquiries	**1 800 700-1033**
Fax line for Depository Services Program	**1 800 889-9734**
E-mail inquiries	**infostats@statcan.ca**
Web site	**www.statcan.ca**

Ordering and subscription information

This product, Catalogue no. 75F0002MIE2004008, is available on Internet free. Users can obtain single issues at: http://www.statcan.ca/cgi-bin/downpub/research.cgi.

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner and in the official language of their choice. To this end, the Agency has developed standards of service which its employees observe in serving its clients. To obtain a copy of these service standards, please contact Statistics Canada toll free at 1 800 263-1136.

Statistics Canada
Income Statistics Division

Income research paper series

Combined-panel longitudinal weighting Survey of Labour and Income Dynamics

1996-2002

Published by authority of the Minister responsible for Statistics Canada

© Minister of Industry, 2004

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission from Licence Services, Marketing Division, Statistics Canada, Ottawa, Ontario, Canada K1A 0T6.

June 2004

Catalogue no. 75F0002MIE2004008

ISSN: 1707-2840

ISBN: 0-662-37553-X

Frequency: Occasional

Ottawa

La version française de cette publication est disponible sur demande (n° 75F0002MIF au catalogue).

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued cooperation and goodwill.

Summary

The Survey of Labour and Income Dynamics (SLID) is a longitudinal survey composed of panels of six years in length. Since the introduction of the second panel in reference year 1996, two panels overlap for periods of three years. Since the beginning of the survey, two types of weights have been produced for each reference year : a longitudinal weight for each panel and a cross-sectional weight which combines data from both panels. The longitudinal weight for one panel allows conducting analyses relating to the population at the time of its selection and that can be carried out over a period of up to six years. However, some SLID data users have expressed the desire to be able to conduct longitudinal analyses using both panels, and thus increasing their precision. The combined panel longitudinal weight has been created to meet this need. It allows doing analyses which refer to the population at the time of the selection of the most recent panel, using individuals from both panels. However, the analyses are limited to the period of three years during which the two panels overlap. This document presents the principles behind the combined panel longitudinal weighting methodology as well as the steps leading to the creation of the weights. These steps are largely inspired from the steps used in the longitudinal weighting of one panel and the cross-sectional weighting. The results obtained with the new weight are briefly evaluated.

Table of Contents

1. Introduction 6
2. SLID Methodology 6
3. Current Longitudinal and Cross-sectional Weights 7
  3.1 Longitudinal weighting 7
  3.2 Cross-sectional weighting 8
4. Combined-Panel Longitudinal Weighting 10
  4.1 The issues 10
  4.2 Definitions 10
  4.3 Target population 11
  4.4 Sample used 11
  4.5 Steps in the combined-panel longitudinal weighting process 13
  4.6 Classification of individuals 14
  4.7 Non-response adjustment 14
  4.8 Migration adjustment 15
  4.9 Combining the panels 15
  4.10 Adjustment for influential values 17
  4.11 Calibration 17
5. Evaluation 17
6. Conclusion 23
References 24

1. Introduction

The Survey of Labour and Income Dynamics (SLID) is a longitudinal panel survey of individuals. Its goal is to measure changes in the economic well-being of individuals and the factors that influence those changes. When it was introduced in the 1993 reference year, the survey was intended to provide longitudinal data, but over the years, its cross-sectional dimension has become just as important. It uses a household sample composed of two panels that are six years in length and overlap for three years.

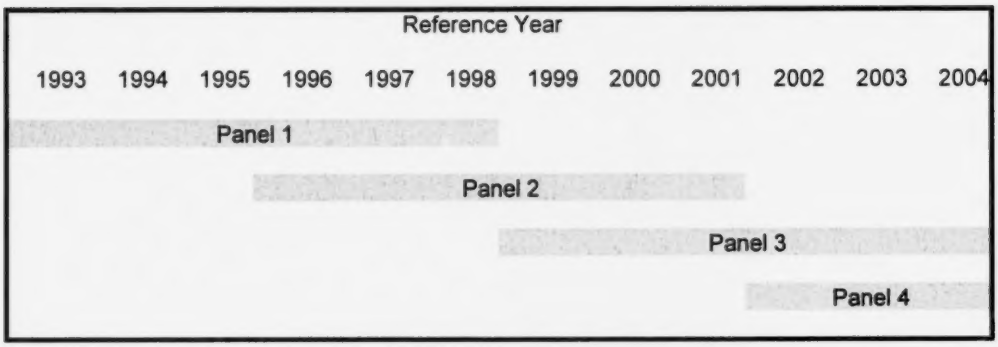
Since the survey's inception, two main types of weights have been produced: a longitudinal weight for each panel, representing the population at the time of selection; and a cross-sectional weight, combining the individuals from both panels for a particular reference year. When longitudinal analyses were conducted, the data for a single panel could be used. Combined-panel longitudinal weighting (CPLW) was developed so that longitudinal studies could use two panels at the same time, thereby doubling the sample size and increasing the precision of the estimates. On the other hand, analyses based on this weighting scheme will be limited to a period of three years.

This paper describes the methodology developed to design and produce combined-panel longitudinal weights. First, SLID's general methodology will be outlined, and a brief description of the longitudinal weighting used for each panel and of the cross-sectional weighting will be provided. Then the various aspects of combined-panel longitudinal weights and the procedure for producing them will be presented. Lastly, there will be a brief assessment of the results obtained with the new weight.

2. SLID Methodology

The Survey of Labour and Income Dynamics is a continuing survey. It is composed of two rotating panels, each six years in length. After the second panel is introduced, there are always two panels at the same time, with each pair of successive panels overlapping for a period of three years. Panel 1 was selected on December 31, 1992, and Panel 2 on December 31, 1995. Since then, a new panel has been selected every three years to replace the older of the two panels, as shown in Figure 1.

Figure 1
Overlap of SLID Panels



The SLID sample consists of about 15,000 households (roughly 40,000 people) for Panel 1 and 17,000 for subsequent panels. The sample is taken from the Labour Force Survey (LFS), whose methodology is described in Singh *et al.* (1990) and Gambino *et al.* (1998). The LFS operates on the basis of six panels; each panel remains in the sample for six months, with one panel replaced each month. The last-stage sampling unit is the dwelling. All members of the households occupying the selected dwellings are included in the LFS sample.

The sample for a SLID panel is composed of households from two outgoing LFS rotation groups in January and February of the first reference year. SLID selects only households that were LFS respondents in January. The final LFS interview serves as the introductory contact for SLID (Lavigne and Michaud, 1998).

Hence, the initial weight for SLID comes directly from LFS and is at the household level. It is used to compute the longitudinal and cross-sectional weights for each wave.

3. Current Longitudinal and Cross-sectional Weights

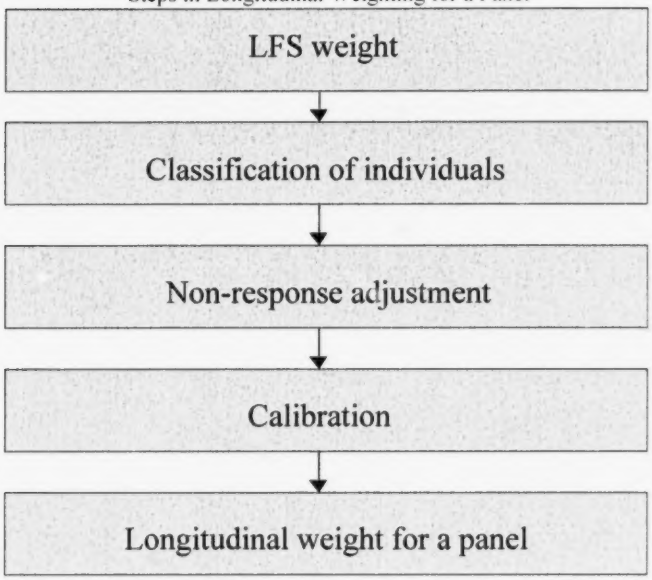
Since it was introduced in 1993, SLID has produced two main types of weights for each reference year: longitudinal weights for each panel and cross-sectional weights for the two panels combined. Combined-panel longitudinal weighting borrows elements from both methodologies. This section contains a brief description of the survey's longitudinal and cross-sectional weights to help the reader understand the development of the new methodology presented in section 4.

3.1 Longitudinal weighting

Before the introduction of combined-panel longitudinal weighting, SLID produced just one type of longitudinal weight. That weight is specific to one of the current panels and represents the population at the time of its selection. It can be used to conduct studies covering the panel's entire six-year lifespan. An overview of the methodology used to produce a panel's longitudinal weight is presented here to make combined-panel longitudinal weighting easier to understand. For a detailed description, see Lévesque and Franklin (2000). In addition, since the steps are also part of the combined-panel longitudinal weighting methodology, they will be described at greater length in section 4.

The target population associated with the longitudinal weight is the population at the time of the panel's selection (December 31, 1992, for Panel 1; December 31, 1995, for Panel 2; and so on). The sample consists of all members of the selected households at the beginning of the panel (longitudinal) and excludes people who joined the households subsequently (cohabitants). The initial weight used for longitudinal weighting is an LFS household weight. However, since the basic longitudinal unit is the individual, SLID's longitudinal weight is at the individual level.

Figure 2
Steps in Longitudinal Weighting for a Panel



Several steps are needed to derive the longitudinal weight for a panel (Figure 2). First, individuals are classified according to whether they are respondents, non-respondents or out of scope (deceased, institutionalized, or outside the 10 provinces). Respondents and out-of-scope individuals will have a non-zero longitudinal weight, and non-respondents will have a weight of zero.

The next step is non-response adjustment. A non-response model is developed, and the weights of respondents are adjusted so that they represent non-respondents as well. Out-of-scope individuals retain their initial weight, thereby representing the portion of the target population that was present at the time of the panel's selection and subsequently left the 10 provinces, entered an institution or died.

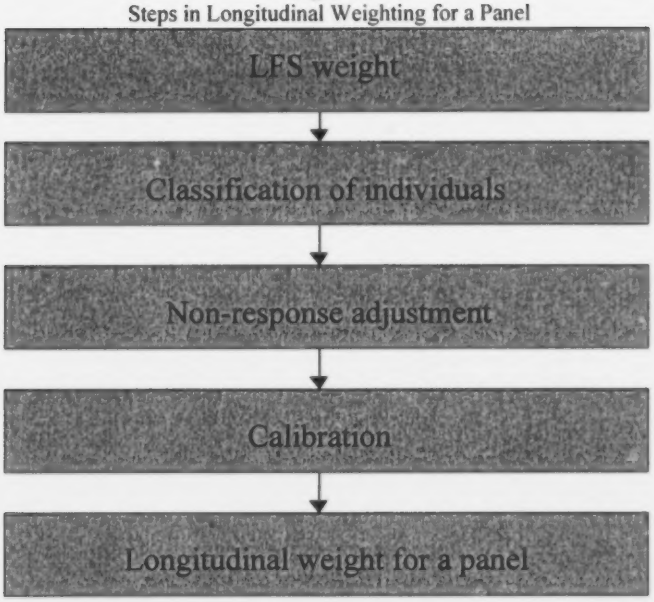
Next, calibration is performed to ensure that certain totals computed with the weights match the population totals derived from other sources. Those totals are, for each province, the number of individuals in each age-sex group, the number of size 1 and 2 economic families, and the number of size 1 and 2 households. They apply to the longitudinal target population, i.e., the population at the time the panel was selected. The result is the final longitudinal weight for the panel. That weight is produced for each reference year. Note that in the near future, calibration will also be based on salary classes (Latouche and Laroche, 2003).

3.2 Cross-sectional weighting

The SLID cross-sectional weight is used to produce estimates for a particular reference year. To that end, the two panels are combined. Individuals who have joined the households of longitudinal persons are referred to as cohabitants and are also part of the cross-sectional sample. A brief overview of the cross-sectional weighting methodology will be presented here. For additional information, see Lévesque and Franklin (2000).

The cross-sectional weight's target population is the population of the 10 provinces on December 31 of the reference year, excluding people living on reserves, in institutions or in military barracks. All longitudinal persons and individuals living in their households (cohabitants) are part of the cross-sectional sample. The

Figure 2
Steps in Longitudinal Weighting for a Panel



Several steps are needed to derive the longitudinal weight for a panel (Figure 2). First, individuals are classified according to whether they are respondents, non-respondents or out of scope (deceased, institutionalized, or outside the 10 provinces). Respondents and out-of-scope individuals will have a non-zero longitudinal weight, and non-respondents will have a weight of zero.

The next step is non-response adjustment. A non-response model is developed, and the weights of respondents are adjusted so that they represent non-respondents as well. Out-of-scope individuals retain their initial weight, thereby representing the portion of the target population that was present at the time of the panel's selection and subsequently left the 10 provinces, entered an institution or died.

Next, calibration is performed to ensure that certain totals computed with the weights match the population totals derived from other sources. Those totals are, for each province, the number of individuals in each age-sex group, the number of size 1 and 2 economic families, and the number of size 1 and 2 households. They apply to the longitudinal target population, i.e., the population at the time the panel was selected. The result is the final longitudinal weight for the panel. That weight is produced for each reference year. Note that in the near future, calibration will also be based on salary classes (Latouche and Laroche, 2003).

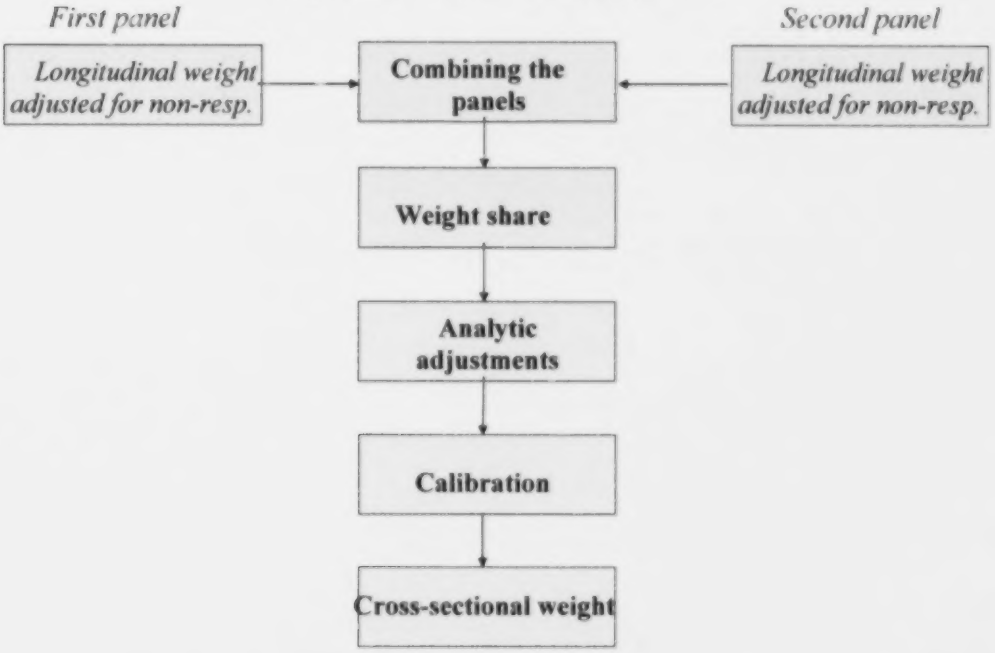
3.2 Cross-sectional weighting

The SLID cross-sectional weight is used to produce estimates for a particular reference year. To that end, the two panels are combined. Individuals who have joined the households of longitudinal persons are referred to as cohabitants and are also part of the cross-sectional sample. A brief overview of the cross-sectional weighting methodology will be presented here. For additional information, see Lévesque and Franklin (2000).

The cross-sectional weight's target population is the population of the 10 provinces on December 31 of the reference year, excluding people living on reserves, in institutions or in military barracks. All longitudinal persons and individuals living in their households (cohabitants) are part of the cross-sectional sample. The

initial cross-sectional weight is the longitudinal weight adjusted for non-response, which represents the population at the time each panel was selected.

Figure 3
Steps in the Cross-sectional Weighting Process



The first step in the cross-sectional weighting process is to combine the samples for the two panels by applying an allocation factor to the non-response-adjusted longitudinal weight. The allocation factor is computed – separately for each province – so as to minimize the variance of a point estimate. No adjustment factor is applied to the weights of Panel 2 individuals who could not have been selected for Panel 1. The panel combination step in the combined-panel longitudinal weighting process, described in section 4.9, is very similar. For more details on panel allocation factors, see Latouche *et al.* (2000) and Merkouris (1999).

The next step in the cross-sectional weighting process is the weight share (Lavallée, 1995). It transfers part of the weight from longitudinal persons to cohabitants who joined their households. Then come the analytic adjustments, one for interprovincial migration and the other for influential values. Lastly, as in the case of longitudinal weighting, calibration is performed against known totals for the reference year (the number of individuals in each age-sex group, the number of size 1 and 2 economic families, and the number of size 1 and 2 households). The result is the final cross-sectional weight. It is produced for each reference year. Note that as in the case of longitudinal weighting, calibration will also be based on salary classes in the near future (Latouche and Laroche, 2003).

4. Combined-Panel Longitudinal Weighting

The purpose of combined-panel longitudinal weighting is to make it possible for longitudinal analyses to use the samples from both panels and thus benefit from the extra precision gained by doubling the sample size. However, since the two panels overlap for only three years, the analyses cannot cover longer periods. Like the other types of weights, a combined-panel longitudinal weight is computed for each reference year.

This section describes in detail the methodology of combined-panel longitudinal weighting. First, the rationale for creating this new weight for SLID will be presented, along with the weight's limitations. This will be followed by information about the target population and the sample. Finally, the steps involved in generating the weight will be outlined.

4.1 The issues

Previously, SLID had two main types of weights: a longitudinal weight for each panel and a cross-sectional weight for the two current panels combined. Longitudinal analyses were, of necessity, based on the sample in just one panel, about 40,000 people. A number of users expressed interest in longitudinal analyses that would be based on both panels. Combining the two panels doubles the sample size and increases the precision of the estimates.

However, since the panels overlap for three years, it is difficult to perform longitudinal studies combining the two panels over a long period. For example, with the weight for Panels 1 and 2, studies can cover only the period from December 31, 1995, to December 31, 1998 (in the case of the weight produced for the 1998 reference year), i.e., from the start of Panel 2 to the end of Panel 1.

4.2 Definitions

To make the text easier to read and understand, the following terms will be used:

First panel:	Denotes the older of the two panels being combined. For example, in the longitudinal weighting of Panels 1 and 2 combined, the first panel will be Panel 1. When Panels 2 and 3 are combined, it will be Panel 2.
Second panel:	Denotes the younger of the two panels being combined. For example, in the longitudinal weighting of Panels 1 and 2 combined, the second panel will be Panel 2. When Panels 2 and 3 are combined, it will be Panel 3.
Panel combination date:	The date associated with the target population of the combined-panel longitudinal weight. The date is also associated with the second panel's target population. For the combination of Panels 1 and 2, the date is December 31, 1995. For Panels 2 and 3, it is December 31, 1998.

In the rest of this paper, the terms "weighting" or "weight" without further qualification are to be understood as referring to combined-panel longitudinal weighting.

4.3 Target population

The target populations of the two panels being combined are different. In both cases, the target population is the Canadian population living in one of the 10 provinces, excluding Indian reserves, military barracks and institutions. However, the two populations are separated by three years. Since the panels begin overlapping as soon as the second panel is introduced, the target population must be identical to the second panel's longitudinal population. For example, when Panels 1 and 2 are combined, the target population will be the same as the Panel 2 target population, i.e., the target population on December 31, 1995.

4.4 Sample used

The combined-panel longitudinal weight must be representative of the target population, and therefore only units that are in that population can be kept in the sample. First-panel longitudinal individuals not in the target population (on the date on which the panels were combined) must be removed from the combined longitudinal sample, even if they may end up in the target population for subsequent years. Those which *are* in the target population will be included in the sample, even if they drop out of the target population in a subsequent year. But it is impossible to identify individuals who are not in the target population if they were non-respondents for the year in which the panels began overlapping. As a result, they have to be considered part of the target population, and the sample, even if they should theoretically be excluded from it. All second-panel longitudinal individuals are part of the target population and the sample, whether they are respondents or not. This does not mean that they will all have a weight greater than zero: non-respondents will have a weight of zero.

For example, when Panels 1 and 2 are combined, the combined longitudinal sample will consist of Panel 1 individuals still living in one of the 10 provinces (excluding institutions, reserves and military barracks) on December 31, 1995, and of all Panel 2 individuals. Panel 1 individuals who were non-respondents in 1995 will have to be included in the target population.

Figure 4
Sample for Combined-Panel Longitudinal Weighting
Example of combining Panels 1 and 2, 1998 reference year

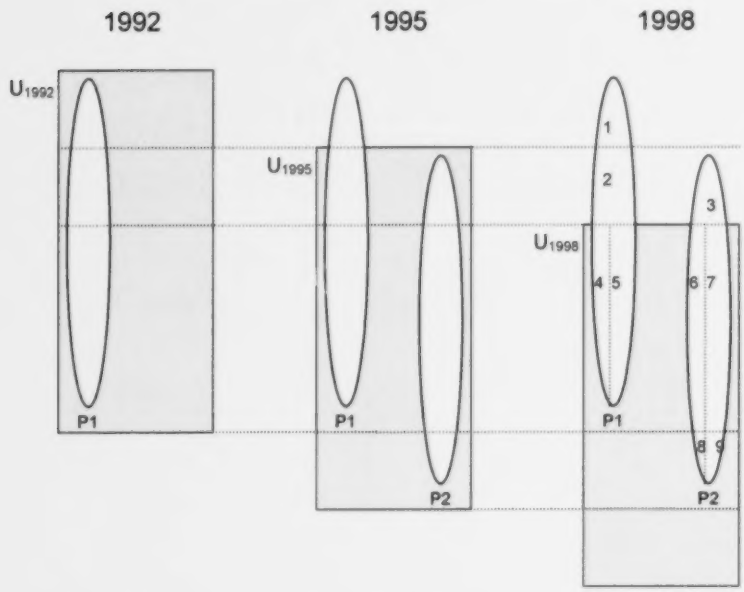


Figure 4 shows the sample used for combining Panels 1 and 2 for the purpose of analyses covering the period from 1995 to 1998 (with the weight produced for the 1998 reference year).

U_{1992} : SLID target population on December 31, 1992 (selection of first panel)
U_{1995} : SLID target population on December 31, 1995 (selection of second panel and panel combination date)
U_{1998} : SLID target population on December 31, 1998 (reference date for the current year)
P1 : Panel 1 sample
P2 : Panel 2 sample

The figures in the various zones of the 1998 samples relate to different categories of individuals:

1: Panel 1 individuals who are not in the target population for combined-panel longitudinal weighting (i.e., on December 31, 1995). These individuals have a weight of zero. Note that the chart has been simplified; in theory, a Panel 1 individual could be out of scope in 1995 and back in the target population in 1998. In that case, he/she would also have a weight of zero.

2 and 3: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting but not in the target population for the 1998 survey. Their weight is greater than zero.

4, 6 and 8: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting (on December 31, 1995) and the target population for the current-year survey (on December 31, 1998) but are non-respondents. These individuals have a weight of zero.

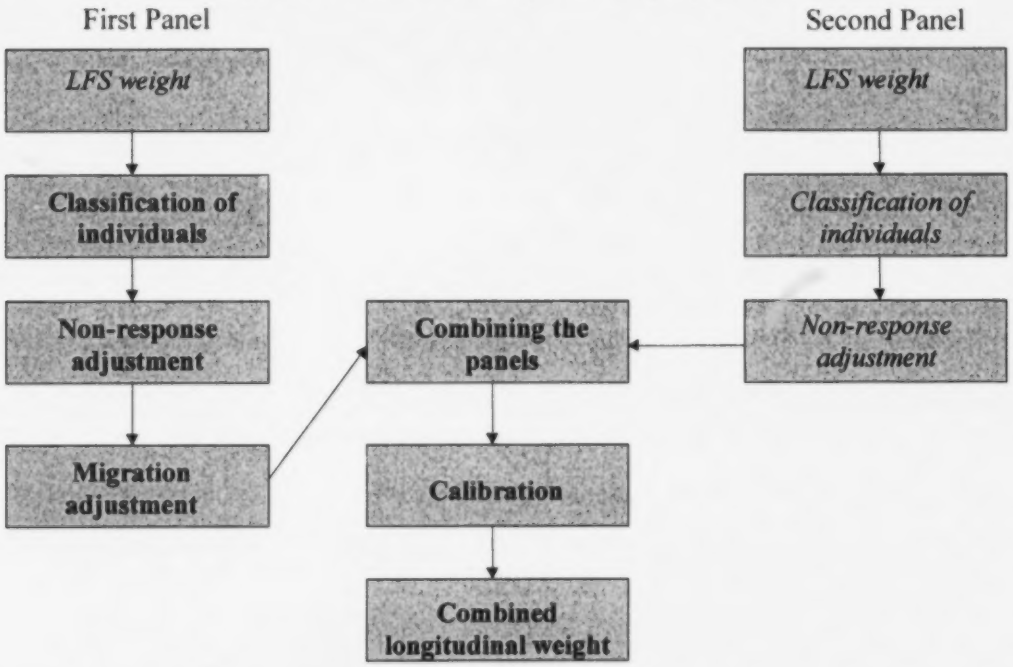
5, 7 and 9: Individuals from the two panels who are in the target population for combined-panel longitudinal weighting (on December 31, 1995) and the target population for the current-year survey (on December 31, 1998) and are respondents. These individuals have a weight greater than zero.

Hence, the sample consists of all individuals from the two panels except the ones in zone 1. Individuals in zones 2, 3, 5, 7 and 9 will have a weight greater than zero.

4.5 Steps in the combined-panel longitudinal weighting process

Producing the combined-panel longitudinal weight is very similar to producing the longitudinal weight for one of the panels. However, the process is more complicated because it involves combining the samples for two panels whose target populations are three years apart.

Figure 5
Steps in the Combined-Panel Longitudinal Weighting Process



Since the two samples were selected independently on dates three years apart, they have to be treated separately at the beginning of the process.

For the second panel, since it already represents the target population on the combination date, the classification of individuals and non-response adjustment will be exactly the same as for the regular longitudinal weighting. Consequently, it will not be necessary to redo these steps, and the panel's non-response-adjusted weight can be used directly in the panel combination step.

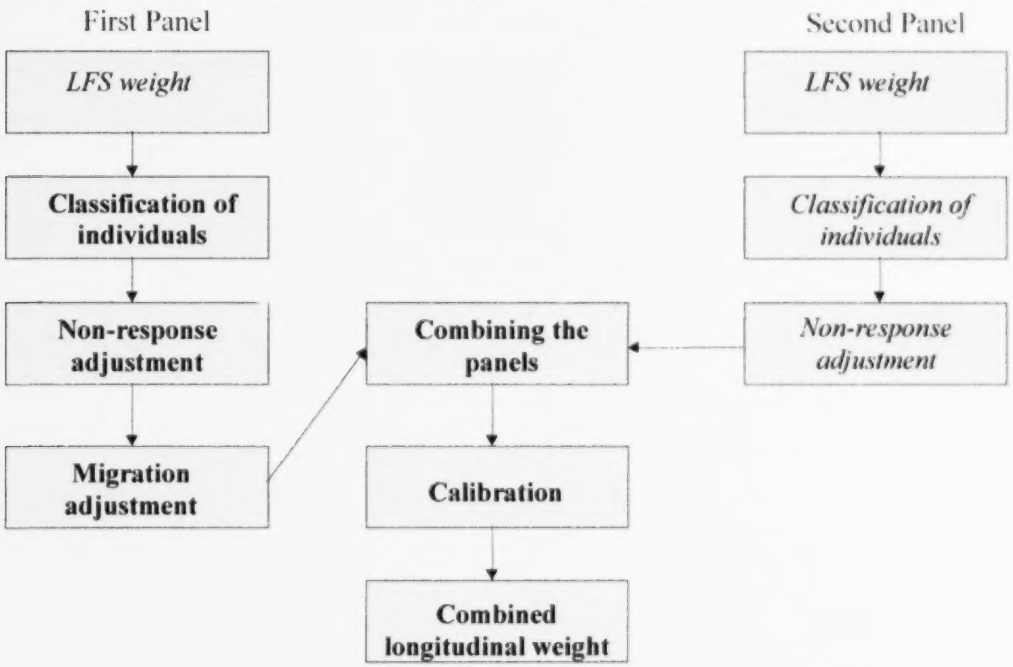
For the first panel, the initial steps have to be redone because the individuals who are not in the target population on panel combination date must be removed from the sample. In other words, the classification of individuals and the non-response adjustment must be redone using only those individuals who are in the target population for combined-panel longitudinal weighting. Then an adjustment is made for interprovincial migration to ensure that individuals who moved from one province to another between the

Hence, the sample consists of all individuals from the two panels except the ones in zone 1. Individuals in zones 2, 3, 5, 7 and 9 will have a weight greater than zero.

4.5 Steps in the combined-panel longitudinal weighting process

Producing the combined-panel longitudinal weight is very similar to producing the longitudinal weight for one of the panels. However, the process is more complicated because it involves combining the samples for two panels whose target populations are three years apart.

Figure 5
Steps in the Combined-Panel Longitudinal Weighting Process



Since the two samples were selected independently on dates three years apart, they have to be treated separately at the beginning of the process.

For the second panel, since it already represents the target population on the combination date, the classification of individuals and non-response adjustment will be exactly the same as for the regular longitudinal weighting. Consequently, it will not be necessary to redo these steps, and the panel's non-response-adjusted weight can be used directly in the panel combination step.

For the first panel, the initial steps have to be redone because the individuals who are not in the target population on panel combination date must be removed from the sample. In other words, the classification of individuals and the non-response adjustment must be redone using only those individuals who are in the target population for combined-panel longitudinal weighting. Then an adjustment is made for interprovincial migration to ensure that individuals who moved from one province to another between the

first panel's sample selection date and the panel combination date do not have a weight that is excessively large relative to other weights.

Next, the panels are combined to produce a weight for all individuals that represents the target population. The last step, as in the case of longitudinal weighting for each panel, is to perform calibration to ensure that the estimates computed with the weights match certain known totals for the population.

4.6 Classification of individuals

As in the case of longitudinal weighting for each panel, the first step in combined-panel longitudinal weighting is the classification of individuals. The two panels are dealt with separately.

For the second panel, the classification will be the same as it was for regular longitudinal weighting. Individuals will be placed in different categories based on whether they are respondents, non-respondents or cross-sectionally out of scope for the reference year.

For the first panel, the process will be identical in all respects except one: individuals who are not in the target population on panel combination date are removed from the sample. In addition, persons who did not respond to the survey for the year in which the second panel was introduced are considered non-respondents for the current reference year, even if they actually responded that year. This is necessary because certain characteristics of respondents on panel combination date are needed for some of the subsequent steps in the weighting process, including calibration.

Table 1
Classification of Individuals
Example: the weighting of Panels 1 and 2 combined, 1998

Categories (from Figure 4)	Description	Number of individuals
	Panel 1	
1	Not in the target population	980
2	Out of scope (cross-sectionally)	1,110
4	Non-respondents	7,493
5	Respondents	30,032
Total		39,615
	Panel 2	
3	Out of scope (cross-sectionally)	1,244
6 and 8	Non-respondents	6,458
7 and 9	Respondents	35,842
Total		43,544
	Panels 1 and 2 combined	
Total		83,159

Table 1 shows the counts for the 1998 reference year. Since about 1,000 individuals from the first panel are not in the target population, the second panel will obviously be more important. Furthermore, because of sample erosion and the slightly more restrictive definition of respondent for the first panel, the number of respondents is higher for the second panel.

4.7 Non-response adjustment

After respondents and non-respondents have been identified, the next step is to adjust for non-response. The only data available for both respondents and non-respondents are the data from the preliminary interview (Panel 1) or the final LFS interview (Panels 2, 3 and 4). Those data are used for modelling. Since the data used for first-panel individuals are three years older than the data used for second-panel

individuals, non-response must be modelled separately for each panel. For Panel 1, for example, even though the target population is the population on December 31, 1995, the adjustment must be based on the data from the preliminary interview (December 31, 1992).

As in the case of longitudinal weighting for each panel, response homogeneity groups (RHGs) are generated by segmentation modelling using chi-square automatic interaction detection (CHAID) (see Kass, 1980), which selects the variable with the highest Pearson chi-square statistic. The weights are then adjusted for non-response in each RHG. The non-response-adjusted weights for the second panel are exactly the same as when the panel is weighted on its own. For the first panel, the adjustment must be redone since individuals who were not in the target population on the panel combination date have been removed from the sample and since the definition of a respondent has an added condition: he/she must also have been a respondent in the year in which the second panel was introduced.

For more details on the SLID non-response adjustment method, see Lévesque and Franklin (2000).

4.8 Migration adjustment

The initial longitudinal weight of first-panel individuals is representative of the target population at the time of its selection, and therefore the weight relates to the province at the time of selection and not to the province of residence on combination date. If in the meantime the individual has moved from a province with a very small sampling fraction (high weight) to a province with a large sampling fraction (low weight), he/she may end up having far too much influence on the new province's estimates.

An interprovincial migration adjustment is made in combined-panel longitudinal weighting for units from the first panel.[1] This is primarily due to the fact that we want in-scope individuals from the first panel to have a longitudinal weight that is representative of the new target population, which is the population at the time the second panel is selected (or on panel combination date).

The adjustment lowers the weight of people who move to a particular province if it is higher than the maximum weight of non-migrants of that same province. The new weight is equal to the 95th percentile in the distribution of non-migrants' weights.

4.9 Combining the panels

At this point, a non-response-adjusted longitudinal weight is available for each panel. For the first panel, that weight is representative of the target population excluding the people who joined it at some point in the three years separating the two panels. For the second panel, the weight already represents the target population. The next step, then, is to combine the two panels to obtain a single panel containing individuals who represent the target population on panel combination date.

Since all first-panel longitudinal individuals who are in scope on panel combination date (zones 2, 4 and 5 in Figure 4) can also be selected for the second panel, their weight is multiplied by a factor p_1 between 0 and 1. The weight of second-panel longitudinal individuals who could have been selected for the first panel (zones 3, 6 and 7 in Figure 4) is multiplied by a factor $p_2 = 1 - p_1$. No factor is applied to second-panel longitudinal persons who were not in the target population when the first panel was selected (zones 8 and 9 in Figure 4). Those individuals were not yet born, outside the 10 provinces, institutionalized or members of the Armed Forces living in barracks when the first panel was selected. Only newborns and international immigrants can be identified. The rest will be considered the same as second-panel longitudinal individuals who could have been selected for the first panel, and their weight will be multiplied by p_2.

1. A similar adjustment is made in SLID's cross-sectional weighting.

The formula used to calculate the panel allocation factors for combined-panel longitudinal weighting is the same as the one used for cross-sectional weighting (Latouche *et al.*, 2000). It minimizes the variance of point estimates made with the sample, before calibration, for all reference years, including the target population (on panel combination date).

Calculation of the panel allocation factors (p) is based on the following formula:

$$p_1 = \frac{n_1}{n_1 + n_2(\frac{d_1}{d_2})}$$

$$p_2 = 1 - p_1$$

An allocation factor is computed for each province. The variables n_1 and n_2 represent the number of individuals aged 16 and over in the two panels who will have a combined longitudinal weight. Only individuals aged 16 and over are considered because there are no income or labour force activity data for children. The variables d_1 and d_2 represent the design effect for the two panels. The design effect is defined as the ratio of the variance obtained with the survey's design to the variance that would be obtained with simple random sampling. The design effect used is the LFS design effect (SLID is an LFS supplement) at the time of the panel's selection; the LFS design effect is associated with the estimated number of people aged 16 and over in the province.

We also considered using allocation factors that minimize the variance of a trend between waves t and t+1. However, that would require more detailed studies, and the increase in precision would probably not be very large in relation to the method used. The following formula is provided for reference purposes:

$$p_1 = \frac{V(\hat{Y}_{t+1,2}) + V(\hat{Y}_{t,2}) - 2COV(\hat{Y}_{t+1,2}, \hat{Y}_{t,2})}{\sum_{j=1}^{2}[V(\hat{Y}_{t+1,j}) + V(\hat{Y}_{t,j}) - 2COV(\hat{Y}_{t+1,j}, \hat{Y}_{t,j})]}$$

where $\hat{Y}_{t,p}$ is the estimate, based on panel p, of a total or average for wave t.

4.10 Adjustment for influential values

The income distribution is asymmetric, with a very long tail for the higher values. As a result, one or more individuals who have both a very large income and a high weight may have an excessive influence on the average income estimates for provinces or smaller domains, and on the associated variance estimate. For that reason, an adjustment for influential values was incorporated in SLID's weighting strategy to lower the weight of such individuals and reduce their influence.

The method used, developed by Tremblay (1998), is applied cross-sectionally. The adjustments are subsequently used for longitudinal weighting as well. The cross-sectional weighting methodology is explained in detail in Lévesque and Franklin (2000).

The influential-value adjustments computed for cross-sectional weighting are applied unchanged during combined-panel longitudinal weighting. However, since this involves longitudinal weighting, the adjustments computed for the preceding years are also applied to the individual's weight. If the individual received an adjustment for more than one reference year, the largest adjustment (i.e., the one with the smallest factor) is applied.

4.11 Calibration

The final step in producing the combined-panel longitudinal weight is calibration on margins. As in the case of longitudinal weighting for each panel, we want the sum of the weights to be equal to certain totals that are known for the target population on panel combination date. Calibration is performed for each province. The control totals used are estimates based on census data: population counts for age-sex groups, the number of size 1 and 2 economic families, and the number of size 1 and 2 households. As in the case of longitudinal weighting for each panel and of cross-sectional weighting, calibration will also be based on salary classes in the near future (Latouche and Laroche, 2003).

The calibration method is generalized regression (GREG).

The resulting weight is the final combined-panel longitudinal weight, which will be used to produce the estimates.

5. Evaluation

Since the combined-panel longitudinal weight always represents the population at the time the panels are combined, it is difficult to evaluate the estimates based on this weight by comparing them with data from other sources. However, since the second panel's longitudinal weight represents the same target population, it is possible to compare the estimates produced with the two weights and the associated variances. The variances are estimated by the bootstrap method (Efron, 1982; Rao and Wu, 1987; Rao, Wu and Yue, 1992). Note that the analysis of the differences between the estimates produced with the two weights can also be interpreted as a comparison of Panels 1 and 2.

Table 2
Place of Residence of Individuals on December 31, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	515 615	522 620	1.36	1.36	1.01	*
Prince Edward Island	124 397	125 431	0.83	1.66	1.42	
Nova Scotia	886 304	881 629	-0.53	1.61	1.25	
New Brunswick	707 558	706 316	-0.18	1.92	1.53	
Quebec	6 928 676	6 913 235	-0.22	0.36	0.29	
Ontario	10 521 030	10 518 677	-0.02	0.38	0.32	
Manitoba	1 003 857	1 006 685	0.28	1.06	0.87	
Saskatchewan	929 087	925 341	-0.40	1.31	1.04	
Alberta	2 637 129	2 652 800	0.59	1.19	0.92	
British Columbia	3 598 511	3 599 722	0.03	0.87	0.69	
In an institution	106 295	103 475	-2.65	10.73	9.17	
Deceased	465 291	466 769	0.32	4.69	3.91	
Total (including persons outside the 10 provinces)	28 733 700	28 733 700	0.00	0.00	0.00	

Table 2 shows the distribution of the longitudinal population by place of residence on December 31, 1998, estimated with the Panel 2 longitudinal weight and with the combined-panel longitudinal weight, the percentage difference between the estimates, and the coefficient of variation for each estimate. The last column indicates whether the difference between the estimates is significant at the 0.05 level; significance is determined from an estimate of the variance of the difference, which is not included in the table. The estimates are very similar, as the coefficients of variation show a significant difference only for Newfoundland. Since the weights are calibrated on the December 31, 1995, population counts, it was to be expected that the estimates would be similar at the provincial level and identical for the total. The estimates produced with the combined-panel longitudinal weight are more precise in every case, though the difference in precision is not very large.

Table 3
Average Income by Province, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	18 697	18 429	-1.44	4.10	2.56	
Prince Edward Island	20 665	21 230	2.73	3.46	2.79	
Nova Scotia	21 329	21 468	0.65	2.82	2.35	
New Brunswick	21 965	21 681	-1.29	2.90	2.09	
Quebec	23 566	23 661	0.40	1.99	1.55	
Ontario	28 359	28 979	2.19	1.73	1.63	*
Manitoba	24 143	23 892	-1.04	3.04	2.32	
Saskatchewan	23 283	23 515	1.00	3.15	2.40	
Alberta	28 352	28 369	0.06	3.23	2.32	
British Columbia	25 955	26 411	1.75	2.44	2.08	
All 10 provinces combined	25 919	26 232	1.21	1.01	0.87	*

Table 3 is similar to Table 2, but it presents estimates of average income by province. The estimates of 1998 average income differ slightly more than the population estimates. For most provinces, the estimate produced with the combined-panel longitudinal weight is slightly higher, although the difference is either not significant or only marginally so. As in the case of Table 2, the use of the combined-panel longitudinal weight increases the precision in all cases, though the increase is quite modest.

Table 4
Proportion of Individuals Below the Low-Income Cut-off, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	13.38%	13.31%	-0.55	12.87	8.58	
Prince Edward Island	9.49%	7.85%	-17.34	20.03	18.67	*
Nova Scotia	14.85%	14.18%	-4.49	9.29	7.82	
New Brunswick	10.03%	10.00%	-0.34	12.98	9.74	
Quebec	16.08%	15.96%	-0.76	7.08	5.49	
Ontario	10.54%	10.03%	-4.83	6.53	5.88	*
Manitoba	14.58%	13.59%	-6.82	11.19	8.93	
Saskatchewan	9.80%	9.29%	-5.20	12.82	10.02	
Alberta	12.10%	11.64%	-3.80	8.15	6.35	
British Columbia	11.77%	11.03%	-6.30	9.08	7.87	
All 10 provinces combined	12.52%	12.07%	-3.58	3.57	2.98	*

Since estimated average income is slightly higher using the combined-panel longitudinal weight, it is no surprise that the estimates of the proportion of individuals below the low-income cut-off would be lower. The differences are appreciable, nearly 0.5% for all provinces combined. The coefficients of variation of the difference also indicate that it is significant for Prince Edward Island, Ontario and all 10 provinces combined. Unfortunately, it is impossible to determine which of the two estimates is more accurate. Nevertheless, the estimates obtained with the combined-panel longitudinal weight are more precise.

Table 5
Average Difference in Total Income Between 1997 and 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
Newfoundland	660	597	-9.49	46.76	33.86	
Prince Edward Island	994	1 048	5.42	31.66	26.96	
Nova Scotia	1 056	1 064	0.76	20.58	17.22	
New Brunswick	1 257	1 238	-1.52	16.37	13.00	
Quebec	1 350	1 338	-0.87	14.08	11.57	
Ontario	1 391	1 706	22.63	14.39	15.17	
Manitoba	1 614	1 531	-5.10	19.65	14.99	
Saskatchewan	874	985	12.76	33.09	21.99	
Alberta	1 680	1 743	3.78	35.94	23.18	
British Columbia	851	804	-5.58	30.94	26.41	
All 10 provinces combined	1 299	1 414	8.83	8.50	8.05	

Table 5 presents estimates of the average increase in personal total income between 1997 and 1998. The large percentage differences between the estimates based on the two types of weights are due to the nature of the variable being estimated. The coefficients of variation are very high, especially for the difference between the estimates. The precision obtained with the combined-panel longitudinal weight is not much better than the precision obtained with the Panel 2 longitudinal weight.

Table 6
Number of Years Below the Low-Income Cut-off Between 1996 and 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight	Significant difference between the estimates Confidence level = 0.05 (*)
0	22 101 268	22 307 533	0.93	0.65	0.53	*
1	2 341 248	2 263 194	-3.33	3.78	3.21	*
2	1 526 475	1 465 215	-4.01	5.43	4.76	*
3	1 883 172	1 816 514	-3.54	5.31	4.44	*
Total	27 852 163	27 852 456	0.00	0.15	0.13	

Table 6 shows the distribution of individuals by the number of years they have spent below the low-income cut-off between 1996 and 1998, the longest possible period for an analysis based on the combined-panel longitudinal weight. The study includes only individuals for whom data are available for all three years. The differences between the estimates produced with the two weights are consistent with the observations regarding previous tables and are significant. The estimates produced with the combined-panel longitudinal weight are slightly more precise.

We might expect the increase in precision yielded by combined-panel longitudinal weighting to be larger for statistics computed for smaller domains. To test this hypothesis, we produced two sets of estimates: the proportion of lone-parent families living below the low-income cut-off, and the average income of immigrants.

Table 7
Proportion of Lone-Parent Families Below the Low-Income Cut-off, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight
Newfoundland	43.22%	44.66%	3.33	30.51	20.39
Prince Edward Island	32.82%	29.29%	-10.74	33.91	30.86
Nova Scotia	70.28%	63.36%	-9.84	9.18	8.59
New Brunswick	43.23%	42.93%	-0.68	14.46	11.82
Quebec	40.05%	39.43%	-1.55	13.93	11.71
Ontario	42.76%	42.23%	-1.24	8.76	7.56
Manitoba	48.43%	47.19%	-2.56	24.03	16.14
Saskatchewan	21.36%	20.67%	-3.24	65.48	47.14
Alberta	31.94%	33.95%	6.27	22.20	15.34
British Columbia	48.29%	39.86%	-17.45	16.55	15.79
All 10 provinces combined	41.86%	40.49%	-3.29	6.01	5.16

Table 7 presents provincial estimates of the proportion of lone-parent families living below the low-income cut-off. Although the estimates are for relatively small domains, the percentage differences between the estimates produced with the two weights are, in most cases, similar to the differences in Table 4. The coefficients of variation are higher, though. It is also clear, especially at the provincial level, that combined-panel longitudinal weighting provides an appreciable increase in precision.

Table 8
Average Income of Immigrants, 1998

	Estimate Panel 2 longitudinal weight	Estimate Combined-panel longitudinal weight	Percentage difference	Coefficient of variation (%) Panel 2 longitudinal weight	Coefficient of variation (%) Combined-panel longitudinal weight
Newfoundland	44 409	40 146	-9.60	38.76	22.08
Prince Edward Island	23 848	26 401	10.70	18.34	11.37
Nova Scotia	23 129	25 015	8.15	14.18	10.87
New Brunswick	32 301	29 365	-9.09	12.37	9.69
Quebec	18 723	19 215	2.63	7.15	5.93
Ontario	27 166	28 043	3.23	2.81	2.59
Manitoba	25 363	23 673	-6.66	6.15	5.01
Saskatchewan	27 325	29 020	6.20	22.17	14.33
Alberta	28 577	28 370	-0.73	7.55	5.47
British Columbia	24 024	24 499	1.98	5.88	4.72
All 10 provinces combined	25 513	26 108	2.33	2.24	1.95

Table 8 presents the average personal income of immigrants and hence is similar to Table 3 except that it covers much smaller domains. The percentage difference between the estimates produced with the two weights is larger for the average income of immigrants than for the average income of all individuals. The increase in precision obtained with the combined-panel longitudinal weight is substantial at the provincial level.

The foregoing tables show that there are differences between the estimates produced with the Panel 2 longitudinal weight and the combined-panel longitudinal weight, even though they are associated with the same target population (the population of the 10 provinces on December 31, 1995) and the samples are partly composed of the same individuals. However, when the sizes of the coefficients of variation are taken into account, the differences are small. They suggest that globally, there is a difference between Panels 1 and 2. However, it is expected that calibration on salary classes, which will be incorporated into the survey's weighting strategy in the near future, will reduce the differences.

Combined-panel longitudinal weighting increases the precision of the estimates in almost every case, but the increase is generally quite modest. On the other hand, most of the estimates presented in the above tables apply to large domains. For smaller subpopulations, there is a much larger gain in precision. Moreover, because Panel 1 is more variable, with rather low allocation factors (between 0.18 and 0.43), the extent to which the combined-panel longitudinal weight can increase the precision relative to the Panel 2 weight is limited. When Panels 2 and 3, which have allocation factors of about 0.5, are combined, the increases in precision are more substantial.

The combined-panel longitudinal weighting methodology described here, including some of the findings in this section, was presented to Statistics Canada's Advisory Committee on Statistical Methods in November 2001 (Latouche, 2001).

6. Conclusion

The combined-panel longitudinal weight was created to provide SLID analysts with the ability to perform longitudinal studies based on two panels. Such studies are more precise, but the longest possible period they can cover is three years, which is the length of the overlap between two successive panels.

The combined-panel methodology is based on SLID's current longitudinal and cross-sectional weighting methodologies. It adjusts the sample of the older panel and combines it with the younger sample to produce a new sample representing the target population. For the older panel, the classification of individuals and the non-response adjustment must be redone. An interprovincial migration adjustment is also performed to ensure that individuals who moved from one province to another during the three-year overlap between the two panels do not have an excessive weight. Then the panels are combined in much the same way as in cross-sectional weighting. Finally, as in the case of longitudinal weighting for each panel, an adjustment is made for influential values and calibration is performed.

To evaluate the estimates produced with the new weight, we compared a number of them for the 1998 reference year with estimates produced with the Panel 2 longitudinal weight, which represents the same target population. We found that for large domains, there are some differences in the estimates, but only a few are significant. There is also an increase in precision with the combined-panel weight. The increase in precision is larger for estimates based on smaller domains.

References

Efron, B. (1982), *The jacknife, the bootstrap and other resampling plans*, Philadelphia: Society for industrial and applied mathematics.

Gambino, J. G., Singh, M. P., Dufour, J., Kennedy, B. and Lindeyer, J. (1998). *Methodology of the Canadian Labour Force Survey*, Statistics Canada. Catalogue 71-526-XPB.

Kass, G.V. (1980). *An exploratory technique for investigating large quantities of categorical data*. Applied Statistics, 29, 119-127.

Latouche, M. (2001). *The Survey of Labour and Income Dynamics: Development of a Combined Panel Longitudinal Weight*. Technical report presented to the Advisory Committee on Statistical Methods, November 2001. Statistics Canada. Internal document.

Latouche, M., Dufour, J. and Merkouris, T. (2000). *Cross-sectional Weighting : Combining Two or More Panels*, Statistics Canada. Catalogue 75F0002MIE-00006.

Latouche, M. and LaRoche, S. (2003), *The Survey of Labour and Income Dynamics Calibration Strategy*, Technical report presented to the Advisory Committee on Statistical Methods, November 2003. Statistics Canada. Internal document.

Lavallée, P. (1995). *Cross-sectional Weighting of the Longitudinal Surveys of Individuals and Households Using the Weight Share Method*, Survey Methodology, vol. 21, no 1, pp. 25-32.

Lavigne, M. and Michaud, S. (1998). *General Aspects of the Survey of Labour and Income Dynamics*, Statistics Canada. Catalogue 75F002MIE-1998005.

Lévesque, I. and Franklin, S. (2000). *Longitudinal and Cross-sectional Weighting of the Survey of Labour and Income Dynamics, 1997 Reference Year*, Statistics Canada. Catalogue 75F0002MIE-00004.

Merkouris, T. (1999). *Cross-sectional Estimation in Multiple Panel Household Surveys*. Statistics Canada. Working paper HSMD 99-004.

Rao, J.N.K. and Wu, C.F.J. (1987). *Resampling Inference with Complex Survey Data*, Journal of the American Statistical Association, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992). *Some Recent Work on Resampling Methods for Complex Surveys*, Survey Methodology, vol.18, n°2, 209-217.

Singh, M. P., Drew, J. D., Gambino, J. G. and Mayda, F. (1990). *Methodology of the Canadian Labour Force Survey*, Statistics Canada. Catalogue 71-526.

Tremblay, J. (1998). *Détection des observations influentes pour l'Enquête sur les finances des consommateurs (EFC) et l'Enquête sur la dynamique du travail et du revenu (EDTR)*, internal document, Statistics Canada.

N° 75F0002MIF au catalogue — N° 008
ISSN: 1707-2867
ISBN: 0-662-77292-X

Document de recherche

Série de documents de recherche - Revenu

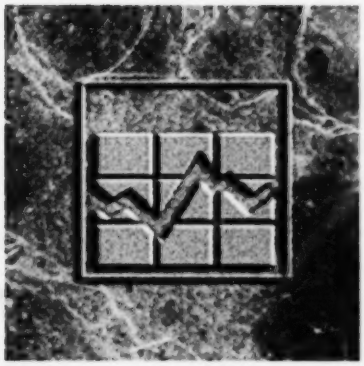
Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu

1996-2002

par Jean-François Naud

Division de la statistique du revenu
Immeuble Jean-Talon, Ottawa, K1A 0T6

Téléphone: 1 613 951-9775



Toutes les opinions émises par l'auteur de ce document ne reflètent pas nécessairement celles de Statistique Canada.

Statistique Canada Statistics Canada

Canada

Comment obtenir d'autres renseignements

Veuillez communiquer avec les Services à la clientèle, Division de la statistique du revenu, Statistique Canada, Ottawa, Ontario, K1A 0T6 ((613) 951-7355; (888) 297-7355: revenu@statcan.ca).

Pour obtenir des renseignements sur l'ensemble des données de Statistique Canada qui sont disponibles, veuillez composer l'un des numéros sans frais suivants. Vous pouvez également communiquer avec nous par courriel ou visiter notre site Web.

Service national de renseignements	**1 800 263-1136**
Service national d'appareils de télécommunications pour les malentendants	**1 800 363-7629**
Renseignements concernant le Programme des bibliothèques de dépôt	**1 800 700-1033**
Télécopieur pour le Progamme des bibliothèques de dépôt	**1 800 889-9734**
Renseignements par courriel	**infostats@statcan.ca**
Site Web	**www.statcan.ca**

Renseignements sur les commandes et les abonnements

Le produit nº 75F0002MIF2004001 au catalogue est gratuit sur Internet. Les utilisateurs sont priés de se rendre à http://www.statcan.ca/cgi-bin/downpub/research_f.cgi.

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois, et ce, dans la langue officielle de leur choix. À cet égard, notre organisme s'est doté de normes de service à la clientèle qui doivent être observées par les employés lorsqu'ils offrent des services à la clientèle. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1 800 263-1136.

Statistique Canada
Division de la statistique du revenu

Série de documents de recherche - Revenu

Pondération longitudinale avec panels combinés, Enquête sur la dynamique du travail et du revenu

1996-2002

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2004

Tous droits réservés. Il est interdit de reproduire ou de transmettre le contenu de la présente publication, sous quelque forme ou par quelque moyen que ce soit, enregistrement sur support magnétique, reproduction électronique, mécanique, photographique, ou autre, ou de l'emmagasiner dans un système de recouvrement, sans l'autorisation écrite préalable des Services de concession des droits de licence, Division du marketing, Statistique Canada, Ottawa, Ontario, Canada K1A 0T6.

Juin 2004

N° 75F0002MIF2004008 au catalogue

ISSN: 1707-2867

ISBN: 0-662-77292-X

Périodicité : hors-série

Ottawa

This publication is available in English upon request (Catalogue no. 75F0002MIE)

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population, les entreprises, les administrations canadiennes et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques précises et actuelles.

Résumé

L'enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale composée de panels d'une durée de 6 ans chacun. Depuis l'introduction du second panel pour l'année de référence 1996, deux panels se chevauchent sur des périodes de 3 ans. Depuis le début de l'enquête, deux types de poids ont été produits à chaque année de référence : un poids longitudinal pour chacun des panels et un poids transversal qui combine les deux panels. Le poids longitudinal de l'un des panels permet d'effectuer des analyses se rapportant à la population au moment de sa sélection et portant sur une période maximale de six ans. Toutefois, des utilisateurs des données de l'EDTR ont manifesté le désir de pouvoir faire des analyses longitudinales en utilisant les deux panels, afin d'en augmenter la précision. Le poids longitudinal avec panels combinés a été créé dans l'objectif de répondre à ce besoin. Il permet de faire, à partir des individus des deux panels, des analyses se rapportant à la population au moment de la sélection du panel le plus récent. Cependant, ces analyses sont limitées à la période de chevauchement des panels, qui est de trois ans. Le présent document présente les principes ayant servi de base à la conception du poids longitudinal avec panels combinés ainsi que les étapes nécessaires à sa création. Ces étapes s'inspirent fortement de celles utilisées dans la pondération longitudinale d'un seul panel et dans la pondération transversale. Les résultats obtenus avec le nouveau poids sont évalués sommairement.

Table des matières

1. Introduction 6
2. Méthodologie de l'enquête 6
3. Pondérations longitudinale et transversale actuelles 7
3.1 Pondération longitudinale 7
3.2 Pondération transversale 8
4. Pondération longitudinale avec panels combinés 10
4.1 Problématique 10
4.2 Définitions 10
4.3 Population cible 11
4.4 Échantillon utilisé 11
4.5 Étapes de la pondération longitudinale avec panels combinés 13
4.6 Classements des individus 14
4.7 Ajustement pour la non-réponse 15
4.8 Ajustement pour la migration 15
4.9 Combinaison des panels 15
4.10 Ajustement pour valeurs influentes 17
4.11 Calage 17
5. Évaluation 17
6. Conclusion 23
Bibliographie 24

1. Introduction

L'Enquête sur la dynamique du travail et du revenu (EDTR) est une enquête longitudinale par panel dont le but est de mesurer les changements au niveau du bien-être économique des individus et les facteurs qui peuvent influencer ces changements. L'enquête, lors de son introduction à l'année de référence 1993, a été d'abord créée dans un but longitudinal, bien que l'aspect transversal soit devenu, au cours des années, tout aussi important. Elle est menée auprès de ménages compris dans deux panels d'une durée de 6 ans et qui se chevauchent sur une période de 3 ans.

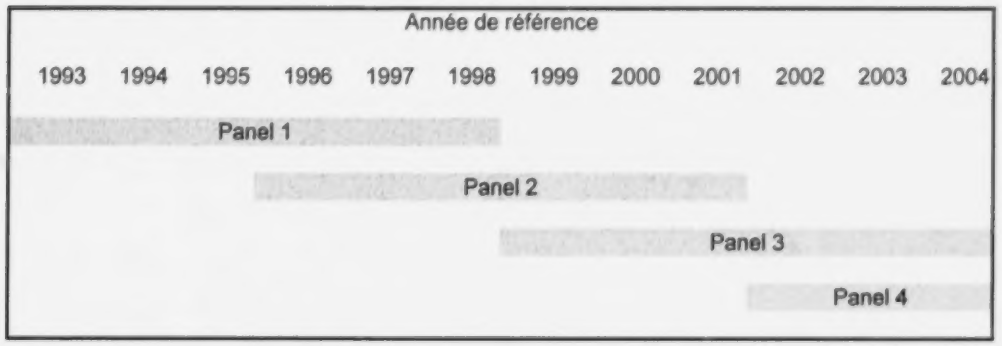
Depuis que l'enquête existe, deux principaux types de poids étaient produits : un poids longitudinal pour chacun des panels représentant la population au moment de leur sélection et un poids transversal, combinant les individus des deux panels, pour une année de référence donnée. Ainsi, lorsque des analyses longitudinales étaient effectuées, les données d'un seul panel pouvaient être utilisées. Le poids longitudinal avec panels combinés (PLPC) a été créé dans le but de permettre des études longitudinales utilisant les deux panels à la fois, doublant ainsi la taille de l'échantillon et augmentant la précision des estimations. Les analyses faites à l'aide de ce poids seront cependant limitées à une période de trois ans.

Ce document présente la méthodologie qui a été développée pour concevoir et produire le poids longitudinal avec panels combinés. D'abord, la méthodologie générale de l'EDTR sera présentée, ainsi qu'une brève description de la pondération longitudinale employée pour chacun des panels et de la pondération transversale. Par la suite, les différents aspects du poids longitudinal avec panels combinés, ainsi que les étapes nécessaires pour le produire, seront décrits. Enfin, on discutera d'une brève évaluation des résultats obtenus avec le nouveau poids.

2. Méthodologie de l'enquête

L'Enquête sur la dynamique du travail et du revenu est une enquête permanente. Elle est composée de deux panels rotatifs d'une durée de 6 ans chacun. À partir de l'introduction du second panel, deux panels coexistent en tout temps, chaque paire de panels successifs se chevauchant pour une durée de trois ans. Le premier panel a été sélectionné en date du 31 décembre 1992, le second panel le 31 décembre 1995. Depuis ce temps, un nouveau panel est sélectionné à tous les trois ans pour remplacer le plus vieux des deux panels existants, comme illustré à la Figure 1.

Figure 1
Chevauchement des panels de l'EDTR



L'échantillon de l'EDTR est composé d'environ 15 000 ménages (environ 40 000 individus) pour le panel 1 et 17 000 ménages pour les suivants. Cet échantillon provient de l'Enquête sur la population active (EPA), dont la méthodologie est décrite dans Singh et coll. (1990) et Gambino et coll. (1998). L'EPA a une base de six panels d'une durée de six mois dont un est remplacé à chaque mois. L'unité d'échantillonnage

au dernier degré de l'EPA est le logement. Tous les individus faisant partie des ménages qui habitent les logements sélectionnés font partie de l'échantillon de l'EPA.

L'échantillon d'un panel de l'EDTR est composé de ménages de deux groupes de renouvellement sortant de l'EPA en janvier et en février de la première année de référence. Les ménages sont sélectionnés uniquement parmi les ménages répondants à l'EPA en janvier. Le premier contact pour introduire l'EDTR est fait lors de la dernière entrevue de l'EPA (Lavigne et Michaud, 1998).

Le poids initial de l'EDTR provient donc directement de l'EPA et est au niveau du ménage. C'est à partir de ce poids que les poids longitudinaux et transversaux sont produits à chacune des vagues.

3. Pondérations longitudinale et transversale actuelles

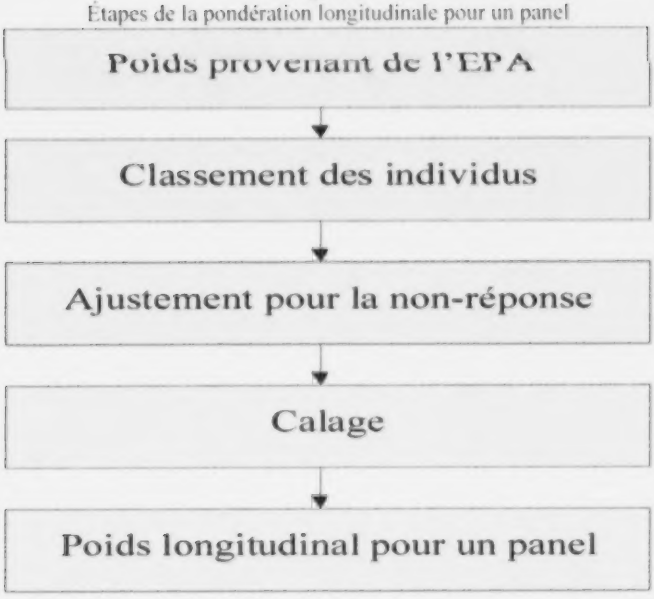
L'EDTR, depuis son introduction en 1993, produit, pour chaque année de référence de l'enquête, deux principaux types de poids : des poids longitudinaux pour chacun des panels et des poids transversaux qui combinent les deux. La pondération longitudinale avec panels combinés emprunte certains éléments à ces deux méthodologies. La présente section décrit brièvement les pondérations longitudinale et transversale de l'enquête afin de permette une meilleure compréhension du développement de la nouvelle méthodologie qui est présentée à la Section 4.

3.1 Pondération longitudinale

Avant l'introduction du poids longitudinal avec panels combinés, l'EDTR produisait un seul type de poids longitudinal. Ce poids est spécifique à l'un des panels en cours et permet de représenter la population au moment de sa sélection. Il permet des analyses sur toute la durée du panel, soit un maximum de 6 ans. Un survol de la méthodologie employée pour produire le poids longitudinal d'un des panels est présenté ici afin de faciliter la compréhension de la pondération longitudinale avec panels combinés. Pour une description détaillée, se référer à Lévesque et Franklin (2000). De plus, puisque chacune des étapes fait aussi partie de la méthodologie de la pondération longitudinale avec panels combinés, elles seront décrites plus amplement dans la section 4 du document.

La population cible liée au poids longitudinal est celle au moment de la sélection du panel (31 décembre 1992 pour le panel 1, 31 décembre 1995 pour le panel 2, etc.). L'échantillon est constitué de tous les individus qui habitaient les ménages sélectionnés en début de panel (longitudinaux) et exclue les personnes qui se sont jointes, au cours des années, aux ménages de ceux-ci (cohabitants). Le poids initial utilisé pour la pondération longitudinale est un poids au niveau du ménage provenant de l'EPA. Par contre, l'unité de base longitudinale étant l'individu, le poids longitudinal de l'EDTR est au niveau individuel.

Figure 2
Étapes de la pondération longitudinale pour un panel



Plusieurs étapes sont nécessaires pour dériver le poids longitudinal pour un panel (figure 2). Tout d'abord, les individus sont classés selon qu'ils sont répondants, non-répondants ou encore hors du champ de l'enquête (mort, en institution, hors des dix provinces). Les répondants et les individus hors du champ de l'enquête auront un poids longitudinal non-nul, alors que les non-répondants auront un poids nul.

L'étape suivante est l'ajustement pour la non-réponse. Un modèle expliquant la non-réponse est créé, et les poids des répondants sont ajustés afin qu'ils représentent aussi les non-répondants. Les individus de l'échantillon qui sont hors du champ de l'enquête conservent leur poids initial et représentent ainsi la partie de la population cible qui était présente au moment de la sélection du panel et qui a par la suite quitté les dix provinces, est entrée en institution ou encore est décédée.

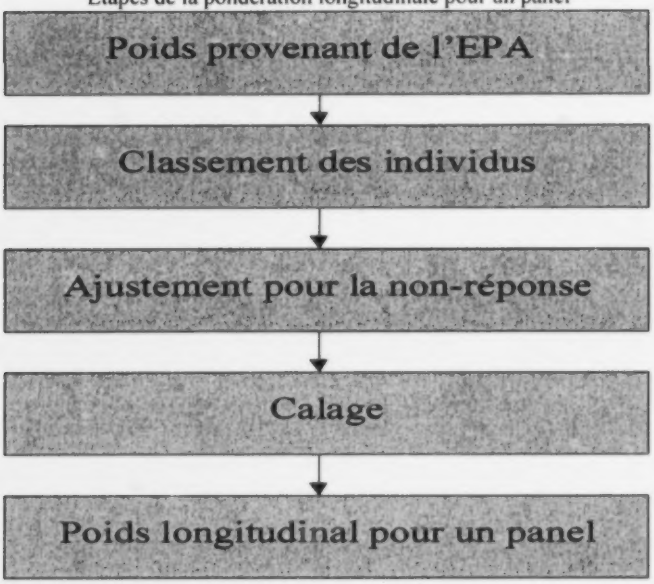
Par la suite, on procède à un calage afin que certains totaux obtenus avec les poids correspondent aux totaux déjà connus pour la population et provenant d'autres sources. Ces totaux sont, pour chaque province, des effectifs d'individus par groupe d'âge et de sexe, le nombre de familles économiques de tailles 1 et 2 ainsi que de ménages de tailles 1 et 2. Ces effectifs sont ceux qui correspondent à la population cible longitudinale, donc au moment de la sélection du panel. On obtient ainsi le poids longitudinal final pour le panel. Ce poids est produit pour chaque année de référence. Il est à noter que, dans un avenir rapproché, le calage se fera également selon les classes de salaire (Latouche et Laroche, 2003).

3.2 Pondération transversale

Le poids transversal de l'EDTR permet de produire des estimations ponctuelles pour une année de référence donnée. Pour ce faire, les deux panels sont combinés. De plus, les individus s'étant joints aux ménages des longitudinaux, les « cohabitants », font aussi partie de l'échantillon transversal. Un bref aperçu de la méthodologie de la pondération transversale sera présenté ici. Pour de plus amples renseignements à ce sujet, le lecteur est prié de se référer à Lévesque et Franklin (2000).

La population cible du poids transversal est la population des dix provinces canadiennes au 31 décembre de l'année de référence, hormis les individus habitant des réserves, des institutions ou des casernements militaires. Tous les longitudinaux de même que les individus habitant dans leur ménage (cohabitants) font

Figure 2
Étapes de la pondération longitudinale pour un panel



Plusieurs étapes sont nécessaires pour dériver le poids longitudinal pour un panel (figure 2). Tout d'abord, les individus sont classés selon qu'ils sont répondants, non-répondants ou encore hors du champ de l'enquête (mort, en institution, hors des dix provinces). Les répondants et les individus hors du champ de l'enquête auront un poids longitudinal non-nul, alors que les non-répondants auront un poids nul.

L'étape suivante est l'ajustement pour la non-réponse. Un modèle expliquant la non-réponse est créé, et les poids des répondants sont ajustés afin qu'ils représentent aussi les non-répondants. Les individus de l'échantillon qui sont hors du champ de l'enquête conservent leur poids initial et représentent ainsi la partie de la population cible qui était présente au moment de la sélection du panel et qui a par la suite quitté les dix provinces, est entrée en institution ou encore est décédée.

Par la suite, on procède à un calage afin que certains totaux obtenus avec les poids correspondent aux totaux déjà connus pour la population et provenant d'autres sources. Ces totaux sont, pour chaque province, des effectifs d'individus par groupe d'âge et de sexe, le nombre de familles économiques de tailles 1 et 2 ainsi que de ménages de tailles 1 et 2. Ces effectifs sont ceux qui correspondent à la population cible longitudinale, donc au moment de la sélection du panel. On obtient ainsi le poids longitudinal final pour le panel. Ce poids est produit pour chaque année de référence. Il est à noter que, dans un avenir rapproché, le calage se fera également selon les classes de salaire (Latouche et Laroche, 2003).

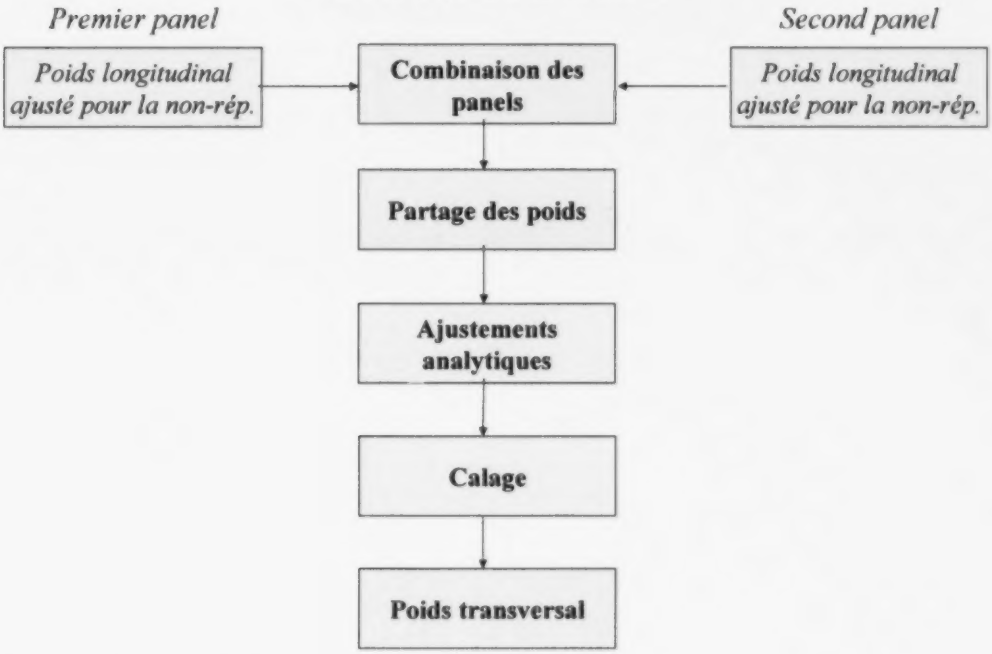
3.2 Pondération transversale

Le poids transversal de l'EDTR permet de produire des estimations ponctuelles pour une année de référence donnée. Pour ce faire, les deux panels sont combinés. De plus, les individus s'étant joints aux ménages des longitudinaux, les « cohabitants », font aussi partie de l'échantillon transversal. Un bref aperçu de la méthodologie de la pondération transversale sera présenté ici. Pour de plus amples renseignements à ce sujet, le lecteur est prié de se référer à Lévesque et Franklin (2000).

La population cible du poids transversal est la population des dix provinces canadiennes au 31 décembre de l'année de référence, hormis les individus habitant des réserves, des institutions ou des casernements militaires. Tous les longitudinaux de même que les individus habitant dans leur ménage (cohabitants) font

partie de l'échantillon transversal. Le poids de départ de la pondération transversal est le poids longitudinal ajusté pour la non-réponse, qui représente la population au moment de la sélection de chacun des panels.

Figure 3
Étapes de la pondération transversale



La première étape de la pondération transversale est de combiner les échantillons des deux panels en appliquant au poids longitudinal ajusté pour la non-réponse un facteur d'allocation. Ce facteur est calculé, pour chaque province, de façon à ce que la variance d'une estimation ponctuelle soit minimisée. Aucun facteur n'est appliqué aux poids des individus du second panel qui n'avaient aucune possibilité de faire partie du premier. L'étape de la combinaison des panels de la pondération longitudinale avec panels combinés, décrite à la Section 4.9, est très semblable. Pour plus de détails sur les facteurs d'allocation des panels, se référer à Latouche et al. (2000) et Merkouris (1999).

L'étape suivante de la pondération transversale est le partage des poids (Lavallée, 1995). Elle permet de transférer une partie du poids des longitudinaux aux cohabitants qui se sont joints à leur ménage. Ensuite, des ajustements analytiques, l'un pour la migration interprovinciale et l'autre pour les valeurs influentes, sont effectués. Finalement, tout comme dans le cas de la pondération longitudinale, on procède à un calage sur des totaux connus pour l'année de référence (effectifs d'individus par groupe d'âge et de sexe, nombre de familles économiques de tailles 1 et 2, nombre de ménages de tailles 1 et 2). Le poids transversal final est ainsi obtenu. Il est produit pour chaque année de référence de l'enquête. Il est à noter que, comme dans le cas de la pondération longitudinale, le calage se fera également, dans un avenir rapproché, selon les classes de salaire (Latouche et Laroche, 2003).

4. Pondération longitudinale avec panels combinés

La pondération longitudinale avec panels combinés a comme objectif de permettre des analyses longitudinales tirant partie des échantillons des deux panels et donc étant plus précises puisque l'on double la taille de l'échantillon. Cependant, étant donné que le chevauchement de deux panels dure seulement 3 ans, les analyses ne pourront pas porter sur des périodes plus longues. Tout comme pour les autres types de poids, un poids longitudinal avec panels combinés est créé pour chaque année de référence de l'enquête.

Dans cette section, la méthodologie détaillée de la pondération longitudinale avec panels combinés sera présentée. D'abord, les raisons qui ont conduit à la création de ce nouveau poids pour l'EDTR, de même que ses limites, seront abordées. Ensuite, il sera question de la population cible ainsi que de l'échantillon utilisé. Puis, chacune des étapes de la création du poids sera décrite.

4.1 Problématique

Auparavant, deux principaux types de poids étaient disponibles à l'EDTR, un poids longitudinal pour chacun des panels ainsi qu'un poids transversal combinant les deux panels en cours. Les analyses longitudinales portaient donc nécessairement sur l'échantillon d'un seul panel, soit environ 40 000 personnes. Plusieurs utilisateurs ont manifesté de l'intérêt pour des analyses longitudinales qui utiliseraient les deux panels. En les combinant, on double la taille de l'échantillon, augmentant ainsi la précision des estimations.

Toutefois, comme le chevauchement des panels s'étale sur 3 ans, il est difficile de faire des études longitudinales combinant les deux panels sur une période plus longue. Ainsi, avec le poids combinant les panels 1 et 2, les analyses ne pourront porter que sur la période allant du 31 décembre 1995 au 31 décembre 1998 (dans le cas du poids produit pour l'année de référence 1998), c'est à dire du début du second panel jusqu'à la fin du premier.

4.2 Définitions

Afin d'alléger le texte et de faciliter la compréhension, les termes suivants seront utilisés :

Premier panel :	désigne le panel le plus vieux parmi les deux qui sont combinés. Par exemple, pour la pondération longitudinale des panels 1 et 2 combinés, il s'agira du panel 1. Pour la combinaison des panels 2 et 3, il s'agira du panel 2.
Second panel :	désigne le panel le plus récent parmi les deux qui sont combinés. Par exemple, pour la pondération longitudinale des panels 1 et 2 combinés, il s'agira du panel 2. Pour la combinaison des panels 2 et 3, il s'agira du panel 3.
Date de combinaison des panels:	date correspondant à la population cible de la pondération longitudinale avec panels combinés. Cette date est aussi celle de la population cible du second panel. Dans le cas de la combinaison des panels 1 et 2, il s'agit du 31 décembre 1995. Pour les panels 2 et 3, il s'agit du 31 décembre 1998.

Dans le reste du document, lorsqu'on traite de « pondération » ou de « poids » sans en préciser le type, c'est toujours dans le contexte de la pondération longitudinale avec panels combinés.

4.3 Population cible

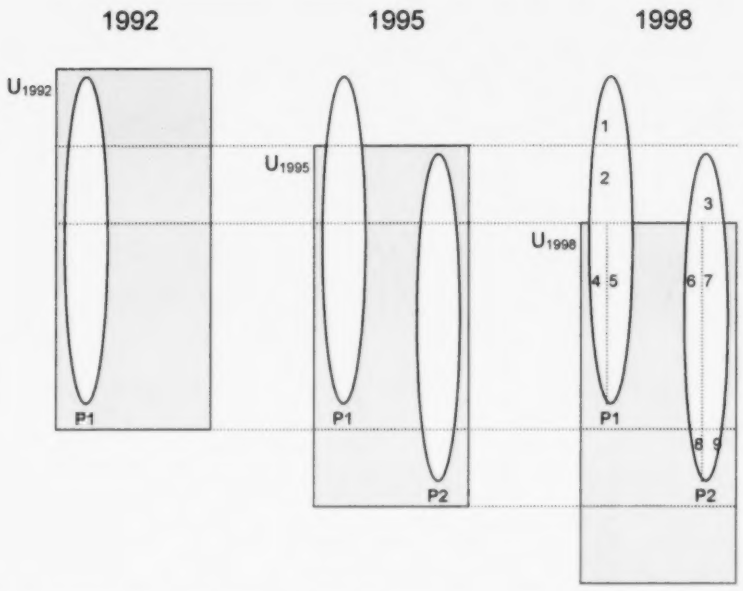
Les populations cibles des deux panels qui sont combinés sont différentes. Dans les deux cas, il s'agit de la population canadienne vivant dans une des 10 provinces, hors des réserves indiennes, des baraques militaires ou des institutions. Cependant, les deux populations sont séparées dans le temps de 3 ans. Comme les panels commencent à se chevaucher au moment de l'introduction du second panel, la population cible devra être identique à la population longitudinale se rapportant à celui-ci. Par exemple, dans le cas de la combinaison des panels 1 et 2, la population cible sera la même que celle du panel 2, et donc en date du 31 décembre 1995.

4.4 Échantillon utilisé

Le poids longitudinal avec panels combinés doit être représentatif de la population cible et donc seules les unités qui en font partie doivent être gardées dans l'échantillon. Parmi les individus longitudinaux du premier panel, ceux qui n'étaient pas dans la population cible (à la date de combinaison des panels) doivent être éliminés de l'échantillon longitudinal combiné, même s'ils reviennent éventuellement dans la population cible des années ultérieures. Les autres, faisant partie de la population cible, feront partie de l'échantillon, et ce même s'ils sortent de la population cible au cours d'une des années suivantes. Il est cependant impossible d'identifier les individus qui sont hors de la population cible si ceux-ci étaient non-répondants à l'enquête pour l'année qui correspond au début du chevauchement des panels. On devra donc les considérer comme faisant partie de la population cible et, par conséquent, de l'échantillon, même s'ils devraient théoriquement en être exclus. Tous les individus longitudinaux du second panel font quant à eux partie de la population cible et de l'échantillon, qu'ils soient répondants ou non. Cela ne veut pas dire qu'ils auront tous un poids plus grand que zéro. Les non-répondants auront un poids nul.

À titre d'exemple, dans le cas de la combinaison des panels 1 et 2, l'échantillon longitudinal combiné sera composé des individus vivants du panel 1 qui résident toujours dans l'une des 10 provinces (excluant les institutions, les réserves et les militaires en casernement) au 31 décembre 1995 et de tous les individus du panel 2. Les non-répondants en 1995 du panel 1 devront être considérés comme faisant partie de la population cible.

Figure 4
Échantillon de la pondération longitudinale avec panels combinés
Exemple de la combinaison des panels 1 et 2, année de référence 1998



La figure 4 illustre l'échantillon qui est utilisé dans le cas de la combinaison des panels 1 et 2, pour des analyses effectuées sur la période 1995-1998 (et donc avec le poids produit pour l'année de référence 1998).

U_{1992} : population cible de l'EDTR au 31 décembre 1992 (sélection du premier panel).
U_{1995} : population cible de l'EDTR au 31 décembre 1995 (sélection du second panel et date de combinaison des panels).
U_{1998} : population cible de l'EDTR au 31 décembre 1998 (date de référence pour l'année en cours).
P1 : échantillon du panel 1.
P2 : échantillon du panel 2.

Les chiffres dans les différentes zones des échantillons des deux panels pour 1998 correspondent à différentes classes d'individus :

1: individus du panel 1 qui ne font pas partie de la population cible de la pondération longitudinale avec panels combinés (i.e. au 31 décembre 1995). Ces individus ont un poids nul. À noter que le graphique a été simplifié ; théoriquement, un individu du panel 1 pourrait être hors-cible en 1995 et de nouveau dans la population cible en 1998. Dans ce cas, il aurait aussi un poids nul.

2 et 3 : individus des deux panels qui font partie de la population cible de la pondération longitudinale avec panels combinés, mais qui ne font pas partie de la population cible de l'enquête en 1998. Ils ont un poids > 0.

4, 6 et 8 : individus des deux panels qui font partie de la population cible de la pondération longitudinale avec panels combinés (au 31 décembre 1995), qui font aussi partie de la population cible de l'enquête pour l'année en cours (au 31 décembre 1998), mais qui sont non-répondants. Ces individus ont un poids nul.

5, 7 et 9 : individus des deux panels qui font partie de la population cible de la pondération longitudinale avec panels combinés (au 31 décembre 1995), qui font aussi partie de la

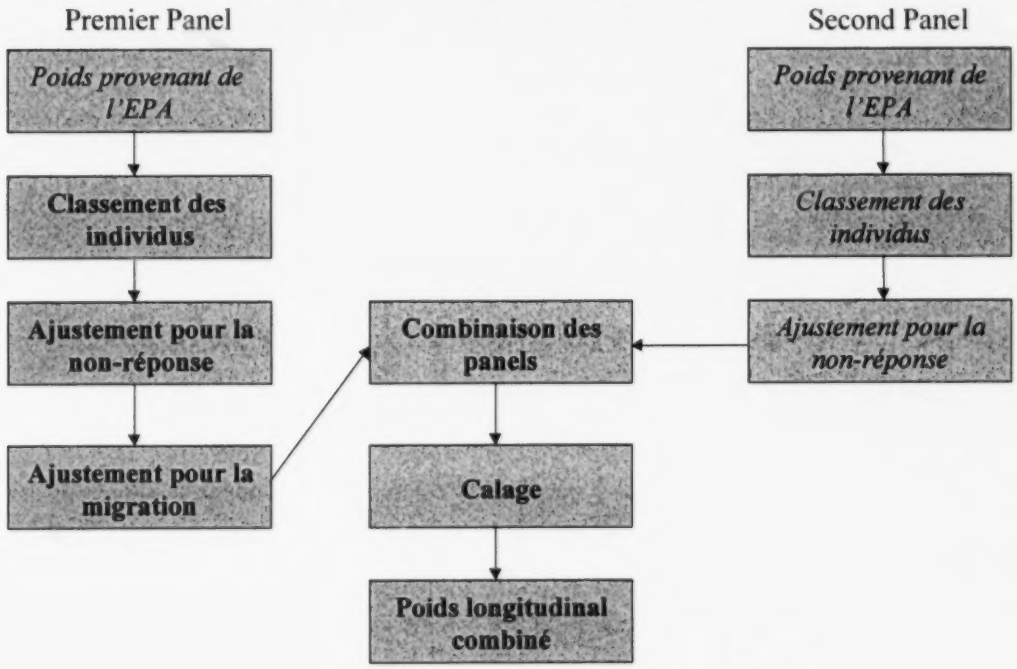
population cible de l'enquête pour l'année en cours (au 31 décembre 1998), et qui sont répondants. Ces individus ont un poids > 0.

L'échantillon se compose donc de tous les individus des deux panels exceptés ceux de la zone 1. Les individus faisant partie des zones 2, 3, 5, 7 et 9 auront un poids > 0.

4.5 Étapes de la pondération longitudinale avec panels combinés

La production du poids longitudinal avec panels combinés est très semblable à celle du poids longitudinal de l'un des deux panels. Elle est cependant plus complexe étant donné qu'on combine les échantillons des deux panels dont les populations cibles sont séparées de 3 ans.

Figure 5
Étapes de la pondération longitudinale avec panels combinés



Puisque les deux échantillons ont été tirés indépendamment et à des dates séparées de 3 ans, il est nécessaire de les traiter individuellement au début du processus.

Pour le second panel, comme il représente déjà la population cible à la date de combinaison, le classement des individus et l'ajustement pour la non-réponse seront identiques à ce qui est fait pour la pondération longitudinale du panel. Il ne sera donc pas nécessaire de refaire ces étapes et le poids après l'ajustement pour la non-réponse de ce panel pourra être directement utilisé pour l'étape de la combinaison des panels.

Pour le premier panel, il faut refaire les premières étapes puisque les individus qui ne font pas partie de la population cible à la date de combinaison des panels doivent être retirés de l'échantillon. Le classement des individus et l'ajustement pour la non-réponse doivent donc être refaits en utilisant uniquement les

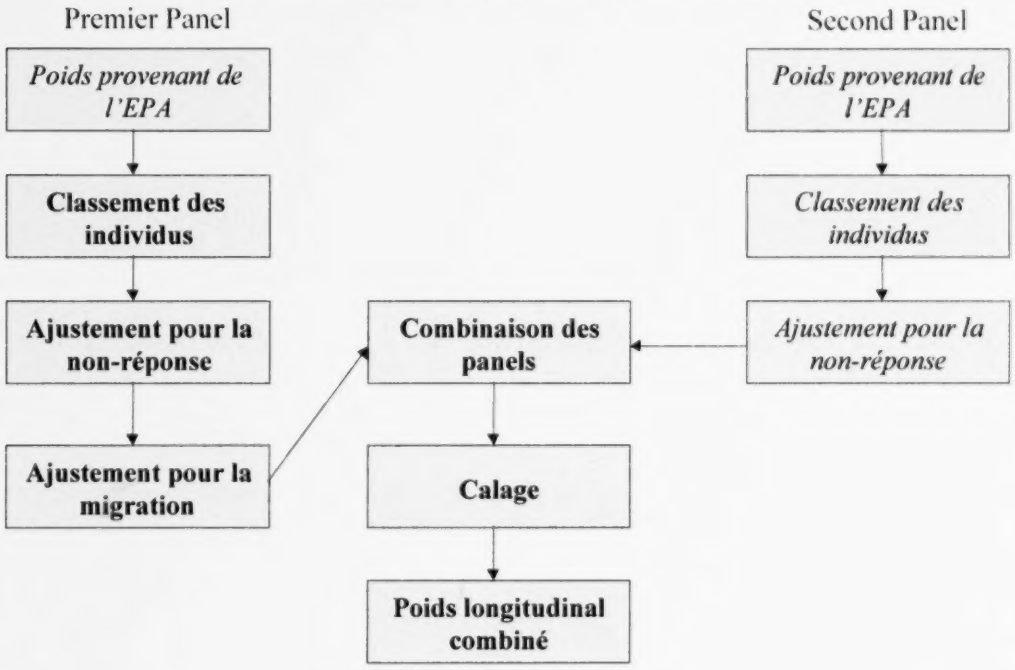
population cible de l'enquête pour l'année en cours (au 31 décembre 1998), et qui sont répondants. Ces individus ont un poids > 0.

L'échantillon se compose donc de tous les individus des deux panels exceptés ceux de la zone 1. Les individus faisant partie des zones 2, 3, 5, 7 et 9 auront un poids > 0.

4.5 Étapes de la pondération longitudinale avec panels combinés

La production du poids longitudinal avec panels combinés est très semblable à celle du poids longitudinal de l'un des deux panels. Elle est cependant plus complexe étant donné qu'on combine les échantillons des deux panels dont les populations cibles sont séparées de 3 ans.

Figure 5
Étapes de la pondération longitudinale avec panels combinés



Puisque les deux échantillons ont été tirés indépendamment et à des dates séparées de 3 ans, il est nécessaire de les traiter individuellement au début du processus.

Pour le second panel, comme il représente déjà la population cible à la date de combinaison, le classement des individus et l'ajustement pour la non-réponse seront identiques à ce qui est fait pour la pondération longitudinale du panel. Il ne sera donc pas nécessaire de refaire ces étapes et le poids après l'ajustement pour la non-réponse de ce panel pourra être directement utilisé pour l'étape de la combinaison des panels.

Pour le premier panel, il faut refaire les premières étapes puisque les individus qui ne font pas partie de la population cible à la date de combinaison des panels doivent être retirés de l'échantillon. Le classement des individus et l'ajustement pour la non-réponse doivent donc être refaits en utilisant uniquement les

individus qui font partie de la population cible de la pondération longitudinale avec panels combinés. Ensuite, un ajustement pour la migration provinciale est fait pour éviter que certains individus ayant migré d'une province à une autre entre la date de sélection de l'échantillon du premier panel et la date de combinaison des panels aient un poids relativement trop grand.

On procède ensuite à la combinaison des panels pour produire un poids pour tous les individus et qui permet de représenter la population cible. Enfin, tout comme pour la pondération longitudinale de chaque panel, on fait un calage pour que les estimations obtenues avec les poids correspondent à certains totaux déjà connus de la population.

4.6 Classements des individus

La première étape de la pondération longitudinale avec panels combinés est, tout comme pour la pondération longitudinale de chacun des panels, le classement des individus. Les deux panels sont traités séparément.

Pour le second panel, le classement sera identique à celui effectué lors de la pondération longitudinale du panel seul. Les individus seront mis dans différentes catégories selon qu'ils sont répondants, non-répondants ou encore non-éligibles transversalement (hors du champ de l'enquête) pour l'année de référence.

Pour le premier panel, le processus sera identique à l'exception du fait que les individus qui ne font pas partie de la population cible (à la date de combinaison des panels) sont retirés de l'échantillon. De plus, les personnes qui n'ont pas répondu à l'enquête pour l'année où le second panel a été introduit sont considérées comme non-répondantes pour l'année de référence en cours, et ce même si elles étaient de fait répondantes pour celle-ci. Ceci est nécessaire étant donné qu'on a besoin de connaître certaines caractéristiques des répondants à la date de combinaison des panels pour certaines étapes ultérieures de la pondération, dont le calage.

Tableau 1
Classement des individus
Exemple de la pondération des panels 1 et 2 combinés, 1998

Classe d'individus (en référence à la figure 4)	Description	Nombre d'individus
Panel 1		
1	Hors de la population cible.	980
2	Non-éligibles (transversalement).	1 110
4	Non-répondants.	7 493
5	Répondants.	30 032
Total		39 615
Panel 2		
3	Non-éligibles (transversalement).	1 244
6 et 8	Non-répondants.	6 458
7 et 9	Répondants.	35 842
Total		43 544
Panels 1 et 2 combinés		
Total		83 159

Le tableau 1 indique les comptes pour l'année de référence 1998. On peut voir qu'étant donné qu'environ mille individus du premier panel ne sont pas dans la population cible, le second panel prendra plus d'importance. De plus, étant donné le phénomène d'érosion ainsi que la définition légèrement plus restrictive de répondant pour le premier panel, le nombre de répondants est plus élevé pour le second panel.

4.7 Ajustement pour la non-réponse

Une fois les répondants et les non-répondants identifiés, c'est l'étape de l'ajustement pour la non-réponse. Les seules données qui sont disponibles à la fois pour les répondants et les non-répondants sont celles qui proviennent de l'entrevue préliminaire (panel 1) ou encore de la dernière entrevue de l'EPA (panels 2, 3 et 4). Ce sont ces données qui sont utilisées pour la modélisation. Ainsi, les données utilisées pour les individus du premier panel sont de trois ans antérieures à celles utilisées pour ceux du second panel. La modélisation de la non-réponse doit donc être faite individuellement pour chacun des panels. Pour le panel 1 par exemple, même si la population cible est celle au 31 décembre 1995, l'ajustement devra être fait à partir des données de l'entrevue préliminaire (31 décembre 1992).

Comme pour la pondération longitudinale de chacun des panels, des groupes homogènes de réponse (GHR) sont créés avec une méthode de modélisation par segmentation utilisant l'algorithme du CHAID (« Chi-Square Automatic Interaction Detection », voir Kass, 1980), qui sélectionne la variable avec la valeur du chi-deux de Pearson la plus élevée. Les poids sont ensuite ajustés pour la non-réponse dans chaque GHR. Les poids ajustés pour la non-réponse pour le second panel sont exactement les mêmes que dans la pondération du panel seul. Pour le premier panel, l'ajustement doit être refait étant donné que les individus qui ne faisaient pas partie de la population cible à la date de combinaison des panels ont été retirés de l'échantillon et que la définition d'un répondant comporte une condition de plus, i.e. la nécessité d'avoir répondu à l'année d'introduction du second panel.

Pour plus de détails sur la méthode d'ajustement pour la non-réponse à l'EDTR, se référer à Lévesque et Franklin (2000).

4.8 Ajustement pour la migration

Le poids longitudinal initial des individus du premier panel est représentatif de la population cible lors de sa sélection et donc le poids se rapporte à la province lors de la sélection et non à la province de résidence à la date de combinaison. Si l'individu a entre temps déménagé d'une province ayant une très petite fraction de sondage (gros poids) vers une province ayant une grande fraction de sondage (petit poids) il risque de devenir beaucoup trop influent pour les estimations de cette nouvelle province.

Un ajustement pour migration interprovinciale est effectué dans le cadre de la pondération longitudinale avec panels combinés pour les unités provenant du premier panel[1]. Ceci essentiellement dû au fait que nous voulons que les individus dans le champ de l'enquête du premier panel aient un poids longitudinal représentatif de la nouvelle population cible, qui est celle lors de la sélection du second panel (ou encore à la date de combinaison des panels).

L'ajustement abaisse le poids des personnes migrantes dans une province si celui-ci est plus élevé que le poids maximum des non-migrants de cette même province. Le nouveau poids est égal au 95^{e} centile de la distribution des poids des non-migrants.

4.9 Combinaison des panels

À ce point, un poids longitudinal ajusté pour la non-réponse est disponible pour chacun des panels. Pour le premier panel, ce poids est représentatif de la population cible, moins les personnes qui se sont jointes à cette population pendant les 3 années qui séparent les deux panels. Pour le second panel, le poids représente déjà la population cible. La prochaine étape est donc de combiner les deux panels pour en obtenir un seul qui réunit les individus représentant la population cible à la date de combinaison des panels.

Puisque tous les individus longitudinaux du premier panel dans le champ de l'enquête à la date de combinaison des panels (zones 2, 4 et 5 de la figure 4) peuvent aussi être choisis lors de la sélection du

1. Un ajustement semblable est effectué pour la pondération transversale de l'EDTR.

second panel, on applique à leur poids un facteur de p_1 compris entre 0 et 1. Un facteur de p_2=1- p_1 est appliqué au poids des individus longitudinaux du second panel qui auraient pu être choisis lors de la sélection du premier panel (zones 3, 6 et 7 de la figure 4). Aucun facteur n'est appliqué aux longitudinaux du second panel pour qui on sait qu'ils ne faisaient pas partie de la population cible lors de la sélection du premier panel (zones 8 et 9 de la figure 4); ce sont les personnes qui n'étaient pas nées, qui étaient en dehors des 10 provinces, qui étaient en institution ou qui des étaient membres des forces armées en casernement au moment de la sélection du premier panel. Seuls les nouveau-nés et les immigrants internationaux peuvent être identifiés; les autres seront considérées aux même titre que les individus longitudinaux du second panel qui auraient pu être choisis lors de la sélection du premier panel et un facteur p_2 sera appliqué à leur poids.

La formule utilisée pour calculer les facteurs d'allocation des panels pour la pondération longitudinale avec panels combinés est la même que celle utilisée pour la pondération transversale (Latouche et al., 2000). Elle a pour effet de minimiser la variance des estimations ponctuelles faites à partir de l'échantillon, avant le calage, pour toutes les années de référence, y compris la population cible (à la date de combinaison des panels).

Le calcul des facteurs d'allocation des panels (p) se base sur la formule suivant :

$$p_1 = \frac{n_1}{n_1 + n_2(\frac{d_1}{d_2})}$$

$$p_2 = 1 - p_1$$

Un facteur d'allocation est calculé pour chacune des provinces. Les variables n_1 et n_2 représentent le nombre d'individus de 16 ans et plus qui auront un poids longitudinal combiné pour chacun des panels. Seuls les individus de 16 ans et plus sont utilisés étant donné qu'il n'y a aucune donnée sur l'activité ou de revenu pour les enfants. Quant à d_1 et d_2, ils représentent l'effet de plan pour chacun des panels. L'effet de plan se définit comme étant le rapport entre la variance obtenue avec le plan de l'enquête et celle qui serait obtenue avec un échantillonnage aléatoire simple. L'effet de plan utilisé est celui de l'EPA (dont l'EDTR est un supplément) au moment de la sélection du panel et associé à l'estimation du nombre de personnes ayant 16 ans et plus dans la province.

On a aussi envisagé d'utiliser des facteurs d'allocation qui minimisent la variance d'une tendance entre les vagues t et t+1. Toutefois, ceci demanderait des études plus approfondies et le gain en précision ne serait probablement pas très important par rapport à la méthode utilisée. La formule suivante est tout de même donnée à titre informatif.

$$p_1 = \frac{V(\hat{Y}_{t+1,2}) + V(\hat{Y}_{t,2}) - 2COV(\hat{Y}_{t+1,2}, \hat{Y}_{t,2})}{\sum_{j=1}^{2}[V(\hat{Y}_{t+1,j}) + V(\hat{Y}_{t,j}) - 2COV(\hat{Y}_{t+1,j}, \hat{Y}_{t,j})]}$$

Où $\hat{Y}_{t,p}$ est l'estimation d'un total ou d'une moyenne à la vague t, faite à partir du panel p.

4.10 Ajustement pour valeurs influentes

La distribution du revenu est asymétrique, étant très étendue pour les valeurs élevées. Ainsi, il peut arriver qu'un ou plusieurs individus, qui ont à la fois un très gros revenu et un poids élevé, aient une trop grande influence sur les estimations de revenu moyen au niveau de la province ou de plus petits domaines, ainsi que sur l'estimation de variance associée. Pour cette raison, un ajustement pour les valeurs influentes a été intégré à la stratégie de pondération de l'EDTR dans le but d'abaisser le poids de ces individus et ainsi de réduire leur influence.

La méthode utilisée, développée par Tremblay (1998), s'effectue au niveau transversal. Les ajustements calculés sont ensuite utilisés également pour la pondération longitudinale. La méthodologie pour la pondération transversale est expliquée en détails dans Lévesque et Franklin (2000).

Les ajustements pour valeurs influentes calculés lors de la pondération transversale sont appliqués tels quels lors de la pondération longitudinale avec panels combinés. Toutefois, comme on est au niveau longitudinal, les ajustements qui ont été calculés pour les années précédentes sont aussi appliqués au poids de l'individu. Dans le cas où l'individu a reçu un ajustement pour plus d'une année de référence, l'ajustement le plus important (i.e. celui dont le facteur est le plus petit) est appliqué.

4.11 Calage

La dernière étape dans la production du poids longitudinal avec panels combinés est celle du calage sur marges. Tout comme pour la pondération longitudinale de chacun des panels, on veut que la somme des poids soient égale a certains totaux déjà connus pour la population cible à la date de combinaison des panels. Le calage est fait par province. Les totaux de contrôle qui sont utilisés sont des estimations faites à partir des données du recensement : effectifs de population par groupes d'age\sexe, nombre de familles économiques de tailles 1 et 2 et nombre de ménages de tailles 1 et 2. Tout comme pour la pondération longitudinale de chacun des panels et la pondération transversale, le calage se fera aussi, dans un avenir rapproché, sur des classes de salaire (Latouche et Laroche, 2003).

La méthode de calage utilisée est celle de la régression généralisée (GREG).

Le poids obtenu est le poids longitudinal avec panels combinés final qui sera utilisé pour produire les estimations.

5. Évaluation

Puisque le poids longitudinal avec panels combinés représente toujours la population au moment de la combinaison des panels, il s'avère difficile d'évaluer les estimations faites à partir de ce poids en comparant avec des données de sources externes. Toutefois, comme le poids longitudinal du second panel représente la même population cible, il est possible de comparer les estimations obtenues avec les deux poids ainsi que les variances associées. Les variances sont estimées selon la méthode d'auto-amorçage ou bootstrap (Efron, 1982; Rao et Wu, 1987; Rao, Wu et Yue, 1992). Il est à noter que l'analyse des différences entre les estimations produites avec les deux poids peut aussi être interprétée comme une comparaison des panels 1 et 2.

Tableau 2
Lieu de résidence des individus au 31 décembre 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés	Différence significative entre les estimations Seuil=0.05 (*)
Terre-Neuve	515 615	522 620	1.36	1.36	1.01	*
Île-du-Prince-Édouard	124 397	125 431	0.83	1.66	1.42	
Nouvelle-Écosse	886 304	881 629	-0.53	1.61	1.25	
Nouveau-Brunswick	707 558	706 316	-0.18	1.92	1.53	
Québec	6 928 676	6 913 235	-0.22	0.36	0.29	
Ontario	10 521 030	10 518 677	-0.02	0.38	0.32	
Manitoba	1 003 857	1 006 685	0.28	1.06	0.87	
Saskatchewan	929 087	925 341	-0.40	1.31	1.04	
Alberta	2 637 129	2 652 800	0.59	1.19	0.92	
Colombie-Britannique	3 598 511	3 599 722	0.03	0.87	0.69	
En institution	106 295	103 475	-2.65	10.73	9.17	
Décédé	465 291	466 769	0.32	4.69	3.91	
Total (incluant individus hors des 10 provinces)	28 733 700	28 733 700	0.00	0.00	0.00	

Le tableau 2 présente la distribution de la population longitudinale selon le lieu de résidence au 31 décembre 1998 estimée avec le poids longitudinal du panel 2 et avec le poids longitudinal avec panels combinés, la différence relative entre ces estimations, de même que les coefficients de variations associés à chacune d'elles. La dernière colonne indique si la différence entre les deux estimations est significative ou non à un seuil de 0,05, ceci étant déterminé à partir d'une estimation de la variance de la différence qui n'est pas incluse dans le tableau. Les estimations sont très semblables, les coefficients de variation n'indiquant pas de différences significatives, sauf dans le cas de Terre-Neuve. Comme les poids sont calés sur les effectifs de population au 31 décembre 1995, on pouvait s'attendre à avoir des estimations semblables au niveau des provinces, et des effectifs identiques pour le total. Les estimations produites avec le poids longitudinal avec panels combinés sont, dans tous les cas, plus précises, quoique cette différence de précision ne soit pas très importante.

Tableau 3
Revenu moyen selon la province, 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés	Différence significative entre les estimations Seuil=0.05 (*)
Terre-Neuve	18 697	18 429	-1.44	4.10	2.56	
Île-du-Prince-Édouard	20 665	21 230	2.73	3.46	2.79	
Nouvelle-Écosse	21 329	21 468	0.65	2.82	2.35	
Nouveau-Brunswick	21 965	21 681	-1.29	2.90	2.09	
Québec	23 566	23 661	0.40	1.99	1.55	
Ontario	28 359	28 979	2.19	1.73	1.63	*
Manitoba	24 143	23 892	-1.04	3.04	2.32	
Saskatchewan	23 283	23 515	1.00	3.15	2.40	
Alberta	28 352	28 369	0.06	3.23	2.32	
Colombie-Britannique	25 955	26 411	1.75	2.44	2.08	
Ensemble des 10 provinces	25 919	26 232	1.21	1.01	0.87	*

Le tableau 3 est semblable au tableau 2, mais présente les estimations de revenu moyen par province. Les estimations du revenu moyen en 1998 faites à partir de chacun des poids diffèrent un peu plus que les estimations de population. Pour la plupart des provinces, on obtient une estimation qui est légèrement plus élevée avec le poids longitudinal avec panels combinés, bien que la différence ne soit pas significative, ou encore très peu. Comme pour le tableau précédent, l'utilisation du poids longitudinal avec panels combinés augmente la précision dans tous les cas, bien que cette augmentation soit assez modeste.

Tableau 4
Proportion d'individus sous le seuil de faible revenu, 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés	Différence significative entre les estimations Seuil=0.05 (*)
Terre-Neuve	13.38%	13.31%	-0.55	12.87	8.58	
Île-du-Prince-Édouard	9.49%	7.85%	-17.34	20.03	18.67	*
Nouvelle-Écosse	14.85%	14.18%	-4.49	9.29	7.82	
Nouveau-Brunswick	10.03%	10.00%	-0.34	12.98	9.74	
Québec	16.08%	15.96%	-0.76	7.08	5.49	
Ontario	10.54%	10.03%	-4.83	6.53	5.88	*
Manitoba	14.58%	13.59%	-6.82	11.19	8.93	
Saskatchewan	9.80%	9.29%	-5.20	12.82	10.02	
Alberta	12.10%	11.64%	-3.80	8.15	6.35	
Colombie-Britannique	11.77%	11.03%	-6.30	9.08	7.87	
Ensemble des 10 provinces	12.52%	12.07%	-3.58	3.57	2.98	*

Comme le revenu moyen estimé est légèrement plus élevé en utilisant le poids longitudinal avec panels combinés, il est normal d'avoir des estimations de la proportion d'individus sous le seuil de faible revenu qui soient inférieures. Ces différences sont quand même assez importantes, près de ½ pourcent pour l'ensemble des provinces. Les coefficients de variation de la différence indique aussi que celle-ci est significative dans le cas de l'Île-du-Prince-Édouard ,de l'Ontario et pour l'ensemble des 10 provinces. Il est malheureusement impossible de déterminer laquelle des deux estimations est la plus proche de la réalité. Toutefois, la précision obtenue avec le poids longitudinal avec panels combinés est meilleure.

Tableau 5
Différence moyenne de revenu total entre 1997 et 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés	Différence significative entre les estimations Seuil=0.05 (*)
Terre-Neuve	660	597	-9.49	46.76	33.86	
Île-du-Prince-Édouard	994	1 048	5.42	31.66	26.96	
Nouvelle-Écosse	1 056	1 064	0.76	20.58	17.22	
Nouveau-Brunswick	1 257	1 238	-1.52	16.37	13.00	
Québec	1 350	1 338	-0.87	14.08	11.57	
Ontario	1 391	1 706	22.63	14.39	15.17	
Manitoba	1 614	1 531	-5.10	19.65	14.99	
Saskatchewan	874	985	12.76	33.09	21.99	
Alberta	1 680	1 743	3.78	35.94	23.18	
Colombie-Britannique	851	804	-5.58	30.94	26.41	
Ensemble des 10 provinces	1 299	1 414	8.83	8.50	8.05	

Le tableau 5 présente les estimations de l'augmentation moyenne du revenu total individuel entre 1997 et 1998. La nature même de la variable estimée explique les écarts relatifs élevés entre les estimations calculées à partir de chacun des types de poids. Les coefficients de variation sont très élevés, en particulier pour ce qui est de la différence entre les estimations. La précision obtenue avec le poids longitudinal avec panels combinés n'est pas de beaucoup supérieure à celle obtenue avec le poids longitudinal du panel 2.

Tableau 6
Nombre d'années sous le seuil de faible revenu entre 1996 et 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés	Différence significative entre les estimations Seuil=0.05 (*)
0	22 101 268	22 307 533	0.93	0.65	0.53	*
1	2 341 248	2 263 194	-3.33	3.78	3.21	*
2	1 526 475	1 465 215	-4.01	5.43	4.76	*
3	1 883 172	1 816 514	-3.54	5.31	4.44	*
Total	27 852 163	27 852 456	0.00	0.15	0.13	

Le tableau 6 contient la distribution des individus selon le nombre d'années passées sur le seuil de faible revenu entre 1996 et 1998, soit la plus longue période possible pour une analyse faite à partir du poids longitudinal avec panels combinés. Seuls les individus dont les données sont disponibles pour les 3 années sont inclus. Les différences entre les estimations produites avec chacun des poids sont cohérentes avec ce qui a été observé dans les tableaux précédents et sont, dans ce cas, significatives. La précision est légèrement supérieure dans le cas des estimations produites avec le poids longitudinal avec panels combinés.

On pourrait s'attendre à ce que le gain en précision apporté par la pondération longitudinale avec panels combinés soit plus important pour des statistiques calculées sur de plus petits domaines. Afin de le vérifier, nous avons produit deux types d'estimation, soient la proportion de familles monoparentales vivant sous le seuil de faible revenu ainsi que le revenu moyen des immigrants.

Tableau 7
Proportion de familles monoparentales sous le seuil de faible revenu, 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés
Terre-Neuve	43.22%	44.66%	3.33	30.51	20.39
Île-du-Prince-Édouard	32.82%	29.29%	-10.74	33.91	30.86
Nouvelle-Écosse	70.28%	63.36%	-9.84	9.18	8.59
Nouveau-Brunswick	43.23%	42.93%	-0.68	14.46	11.82
Québec	40.05%	39.43%	-1.55	13.93	11.71
Ontario	42.76%	42.23%	-1.24	8.76	7.56
Manitoba	48.43%	47.19%	-2.56	24.03	16.14
Saskatchewan	21.36%	20.67%	-3.24	65.48	47.14
Alberta	31.94%	33.95%	6.27	22.20	15.34
Colombie-Britannique	48.29%	39.86%	-17.45	16.55	15.79
Ensemble des 10 provinces	41.86%	40.49%	-3.29	6.01	5.16

Le tableau 7 présente les estimations de la proportion de familles monoparentales sous le seuil de faible revenu, pour chacune des provinces. Bien que les estimations soient faites sur des domaines relativement petits, les différences relatives entre celles obtenues à partir du poids longitudinal du panel 2 et celles obtenues à partir du poids longitudinal avec panels combinés sont, dans la plupart des cas, semblables à celles du tableau 4. Les coefficients de variations sont cependant plus élevés. Il est aussi clair, surtout à l'échelle provinciale, que la pondération longitudinale avec panels combinés apporte un gain important en précision.

Tableau 8
Revenu moyen des immigrants, 1998

	Estimation poids longitudinal panel 2	Estimation poids longitudinal avec panels combinés	Différence relative (%)	Coefficient de variation (%) – poids longitudinal panel 2	Coefficient de variation (%) – poids longitudinal avec panels combinés
Terre-Neuve	44 409	40 146	-9.60	38.76	22.08
Île-du-Prince-Édouard	23 848	26 401	10.70	18.34	11.37
Nouvelle-Écosse	23 129	25 015	8.15	14.18	10.87
Nouveau-Brunswick	32 301	29 365	-9.09	12.37	9.69
Québec	18 723	19 215	2.63	7.15	5.93
Ontario	27 166	28 043	3.23	2.81	2.59
Manitoba	25 363	23 673	-6.66	6.15	5.01
Saskatchewan	27 325	29 020	6.20	22.17	14.33
Alberta	28 577	28 370	-0.73	7.55	5.47
Colombie-Britannique	24 024	24 499	1.98	5.88	4.72
Ensemble des 10 provinces	25 513	26 108	2.33	2.24	1.95

Le tableau 8 présente le revenu individuel moyen des immigrants, et donc est l'équivalent du tableau 3 mais pour des domaines beaucoup plus petits. La différence relative entre les estimations produites à partir du poids longitudinal du panel 2 et celles produites à partir du poids longitudinal avec panels combinés est plus grande pour le revenu moyen des immigrants que pour le revenu moyen de l'ensemble des individus.

Le gain en précision obtenu avec le poids longitudinal avec panels combinés est important à l'échelle provinciale.

Les résultats des tableaux précédents indiquent qu'il existe des différences entre les estimations produites avec le poids longitudinal du panel 2 et celles produites avec le poids longitudinal avec panels combinés, bien qu'ils soient liés à la même population cible, celle des 10 provinces au 31 décembre 1995 et que les échantillons soient composés, en partie, des mêmes individus. Toutefois, quand on prend en considération les tailles des coefficients de variation, ces différences demeurent légères. Elles suggèrent tout de même qu'il existe, de façon globale, une différence entre les panels 1 et 2. On peut cependant s'attendre à ce que le calage sur les classes de salaire, qui sera incorporé prochainement à la stratégie de pondération de l'enquête, vienne atténuer ces différences.

La pondération longitudinale avec panels combinés apporte un gain de précision pour la presque totalité des estimations, mais ce gain est généralement assez modeste. Par contre, la plupart des estimations présentées dans les tableaux ci-dessus sont produites sur de grands domaines. Pour des sous-populations plus restreintes, le gain en précision est beaucoup plus important. De plus, le panel 1 étant plus variable, avec des facteurs d'allocation assez bas (entre 0,18 et 0,43), cela limite l'augmentation de la précision pour le poids longitudinal avec panels combinés par rapport au poids du panel 2. Lors de la combinaison des panels 2 et 3, dont les facteurs d'allocation sont environ de 0,5, les gains en précision sont plus appréciables.

La méthodologie de la pondération longitudinale avec panels combinés, présentée dans ce document, incluant certains des résultats de la présente section, a été présentée au Comité consultatif sur les méthodes statistiques de Statistique Canada, en novembre 2001 (Latouche, 2001).

6. Conclusion

Le poids longitudinal avec panels combinés a été créé dans l'objectif de fournir aux analystes de l'Enquête sur la dynamique du travail et du revenu la possibilité de faire des analyses longitudinales tirant profit des deux panels. Ces analyses sont donc plus précises, bien qu'elles ne puissent être effectuées que sur une période maximale de trois ans correspondant au chevauchement de deux panels successifs.

La méthodologie de la pondération longitudinale s'inspire des pondérations longitudinales et transversales actuelles de l'EDTR. Elle permet d'ajuster l'échantillon du panel le plus âgé et de le combiner au plus récent de façon à ce que le nouvel échantillon représente la population cible. Pour ce faire, il faut refaire la classification des individus du panel le plus âgé ainsi que l'ajustement pour la non-réponse. On fait aussi un ajustement pour la migration provinciale afin d'éviter que des individus ayant changé de province pendant les trois années séparant la sélection des deux panels aient un poids trop important. On procède ensuite à la combinaison des panels de manière similaire à ce qui est fait pour le transversal. Puis, comme lors de la pondération longitudinale de chacun des panels, on effectue un ajustement pour les valeurs influentes et un calage.

Afin d'évaluer les estimations produites avec le nouveau poids, nous en avons comparé certaines, pour l'année de référence 1998, avec celles produites à partir du poids longitudinal du panel 2, qui représente la même population cible. Nous avons pu en tirer que, sur de grands domaines, certaines différences existent dans les estimations, bien que seulement quelques-unes soient significatives. De plus, il y a un certain gain de précision en utilisant le poids avec panels combinés. Ce gain est plus important lors d'estimations sur des domaines plus restreints.

Bibliographie

Efron, B. (1982), *The jacknife, the bootstrap and other resampling plans*, Philadelphia: Society for industrial and applied mathematics.

Gambino, J. G., Singh, M. P., Dufour, J., Kennedy, B. et Lindeyer, J. (1998). *Méthodologie de l'enquête sur la population active du Canada*, Statistique Canada. Catalogue 71-526-XPB.

Kass, G.V. (1980). *An exploratory technique for investigating large quantities of categorical data*. Applied Statistics, 29, 119-127.

Latouche, M. (2001). The Survey of Labour and Income Dynamics: Development of a Combined Panel Longitudinal Weight. Rapport technique présenté au Comité consultatif sur les méthodes statistiques, novembre 2001. Statistique Canada. Document interne.

Latouche, M., Dufour, J. et Merkouris, T. (2000). *Pondération transversale : combinaison de deux panels ou plus*, Statistique Canada. Catalogue 75F0002MIF-00006.

Latouche, M. et LaRoche, S. (2003), The Survey of Labour and Income Dynamics Calibration Strategy, Rapport technique présenté au Comité consultatif sur les méthodes statistiques, novembre 2003. Statistique Canada. Document interne.

Lavallée, P. (1995). *Pondération transversale des enquêtes longitudinales menées auprès des individus et des ménages à l'aide de la méthode du partage des poids*, Techniques d'enquête, vol. 21, no 1, pp. 27-35.

Lavigne, M. et Michaud, S. (1998). *Aspects généraux de l'Enquête sur la dynamique du travail et du revenu*, Statistique Canada. Catalogue 75F002MIF-1998005.

Lévesque, I. et Franklin, S. (2000). *Pondération longitudinale et transversale de l'Enquête sur la dynamique du travail et du revenu. Année de référence 1997*, Statistique Canada. Catalogue 75F0002MIF-00004.

Merkouris, T. (1999). *L'estimation transversale dans les enquêtes-ménages à panels multiples*. Statistique Canada. Document de travail HSMD 99-004.

Rao, J.N.K. et Wu, C.F.J. (1987). *Resampling Inference with Complex Survey Data*, Journal of the American Statistical Association, 83, 231-241.

Rao, J.N.K., Wu, C.F.J. et Yue, K. (1992). *Quelques travaux récents sur les méthodes de rééchantillonnage applicables aux enquêtes complexes*, Techniques d'enquête, vol.18, n°2, 225-234.

Singh, M. P., Drew, J. D., Gambino, J. G. et Mayda, F. (1990). *Méthodologie de l'enquête sur la population active du Canada*, Statistique Canada. Catalogue 71-526.

Tremblay, J. (1998). *Détection des observations influentes pour l'Enquête sur les finances des consommateurs (EFC) et l'Enquête sur la dynamique du travail et du revenu (EDTR)*, document interne, Statistique Canada.